A molecular analysis of hyaluronate lyase production in *Streptococcus pneumoniae*

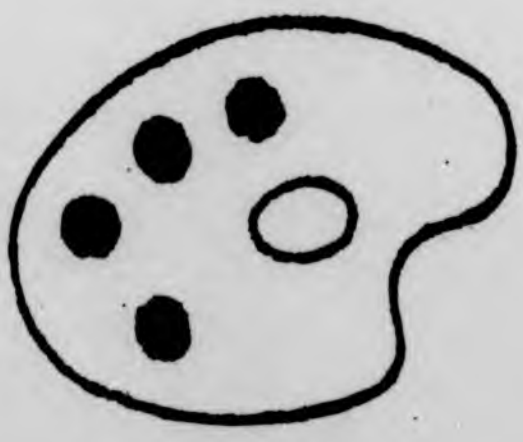
Neil Christopher Doherty

A thesis submitted for the degree of Doctor of Philosophy

Department of Biological Sciences
University of Warwick

September 2000

Numerous Originals in Colour



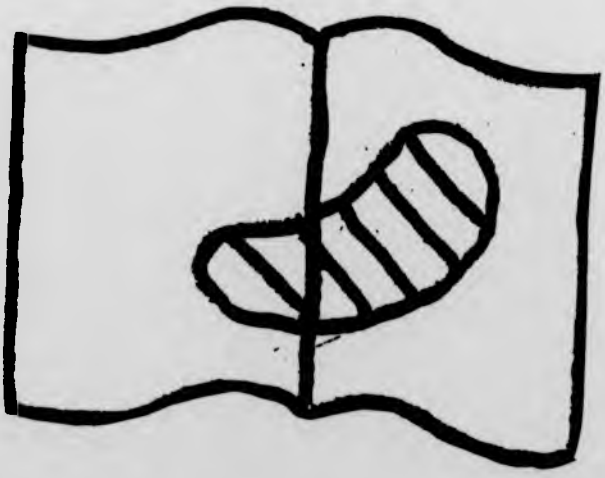


Table of Contents

List of Figures

List of Tables

Chapter 2 **Materials and Methods**

Acknowledgements

I would like to thank my supervisor, Professor Chris Dowson for allowing me the opportunity to carry out the work presented in this thesis, and for all his help over the years.

I would also like to thank the past and present members of the group for (variously) advice, opinion, constructive criticism, destructive criticism, amusement and bewilderment. Frequently concurrently.

I would especially like to thank Dr. Adrian Whatmore. A busy man who is always willing to help.

I would like to thank Richard Waite for his help with the biofilm work. I am also extremely grateful for assistance and advice from Dr. Karen Homer and Professor David Beighton at King's College School of Medicine and Dentistry.

This thesis would not have come to pass without the constant encouragement and (frequently financial) support of my Mum, Dad, Jonathan and Marivel.

Thanks also to Helen for her constant support and encouragement, Nick and Rich for revelry and associated hangovers. Also, I would especially like to thank Joan and Barry for restoring my faith in human nature!

Finally I would like to thank Mr. Andrew Johnson for first making it apparent that the pursuit of science is something to be enjoyed, not endured.

Declaration

I hereby declare that the work presented in this thesis is my own work unless otherwise stated in the text or figure legends. This work has not been submitted to any other institution or for any previous degree. All information sources are acknowledged.

Neil Doherty

Abstract

Streptococcus pneumoniae is a global cause of mortality and morbidity. Its genetic plasticity renders it responsive to the selective pressures of the host environment and antimicrobial chemotherapy.

Hyaluronidase (hyaluronate lyase) is an important protein virulence determinant of this organism, which has previously been described as a spreading factor, facilitating invasion of host tissues during episodes of invasive disease by degrading the hyaluronic acid component of the host extracellular matrix.

The available protein sequence indicates that hyaluronate lyase is exposed to the extracellular environment. Selection to drive variation at the *hyl* locus by the host immune response was investigated using high-resolution restriction analysis (HRRA) and direct sequence analysis. This revealed that variation of *hyl* is low, however, there is statistically significant evidence that variation at *hyl* is mediated by homologous recombination. The low overall divergence suggests that recombination is restricted to the intra-species level.

Phenotypic analysis of hyaluronate lyase activity did not reveal statistically significant associations with site of isolation or genotype. Hyaluronate lyase was shown to be inducible by hyaluronate.

Mutagenesis of the *hyl* locus reveals that other pneumococcal enzymes do not functionally complement loss of the enzyme.

Hyaluronic acid (HA) is able to support growth of some strains of *S. pneumoniae*, and hyaluronate lyase has been shown to be required for growth. Pathways for the utilisation of HA derived metabolites are proposed in the light of available pneumococcal genome data, and published observations in other systems.

The previously published *hyl* gene sequence is thought to be incomplete. Elements upstream of *hyl* have been isolated and show putative strong promoter and cleavable N-terminal signal sequence.

Abbreviations

Abbreviations frequently used in this thesis are listed below in alphabetical order. Abbreviations omitted from this list include strains of bacteria, accepted gene or phenotype designations, names of plasmids and restriction enzymes.

A	Adenine (base)
A/Abs.	Absorbance
ab	Antibody
anova	Analysis of variance
BHI	Brain-Heart Infusion (media)
bp	base pair
BSA	Bovine Serum Albumin
C	Cytosine (base)
CAT	Chloramphenicol Acetyltransferase
CL	Confidence Limits
CM	C-medium
CMHA	C-medium complemented with hyaluronate
CSP	Competence stimulating peptide
ΔUA GlcNAc	N-acetylhyalobiuronic acid
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ET	Electrophoretic type
et al.	*et alia* – and others
G	Guanine (base)
GlcNAc	N-acetylglucosamine
GlcUA	Glucuronic acid
h	Hours
HA	Hyaluronic acid
HRRA	High resolution restriction analysis
Ig	Immunoglobulin
iPCR	Inverse polymerase chain reaction
kb	Kilobase
kbl	Kilobase ladder (electrophoretic marker)
kDA	KiloDalton
l	Litre
LB	Luria Bertani (media)
m	minute(s)
M	Molar
µg	microgramme
µl	microlitre
ml	mililitre
mM	milimolar

nm	nanometres
O.D.	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PVD	Protein virulence determinant
r.p.m	Revolutions per minute
RT	Restriction type
S.D.	Standard deviation
STM	Signature tagged mutagenesis
T	Thymine (base)
TA	Thymine/Adenine overhangs
Tris	Tris(hydroxymethyl)aminomethane
U.V.	Ultra violet
v/v	volume to volume
w/v	weight to volume

To my family

1.1 - The pneumococcus from a historical perspective

"No other disease kills from one fourth to one half of all persons attacked...so fatal is it, that to die of pneumonia in this country is said to be the natural end of elderly people".

Those were the words of Dr. William Osler circa1900, one of the foremost American physicians of his time (Golden, 1992). *Streptococcus pneumoniae* (or the pneumococcus) is an organism responsible for widespread global mortality and morbidity, yet investigations involving this pathogen have helped lay the foundation stones of molecular biology and immunology. When addressing the balance between man and disease, the place of *S. pneumoniae* is firmly rooted in scientific history.

In the nineteenth century, immediately prior to characterisation of the pneumococcus, the spread of infections by this organism was exacerbated by socio-economic conditions prevailing over large sections of society, both within Britain, and globally. To the 'lower' classes, high local population densities, poor housing and sanitation (and the lack of an understanding of a need for such) and immunocompromisation imposed by dietary constraints all contributed to high mortality resulting from a plethora of infectious diseases.

However, pneumococcal infection was by no means restricted to the socially or financially disadvantaged. Individuals could be struck down in any walk of life. This was a key feature described at the beginning of the nineteenth century by Laennec, who also noted 'variable but significant mortality' (Foster, 1970).

In terms of human suffering, it is now understood that *S. pneumoniae* is the causative agent of the severest type of pneumonia, meningitis, bacteraemia, septicaemia, and endocarditis, as well as the non-life threatening conditions otitis media and arthritis. The route to the isolation and characterisation of the pneumococcus was conducted through investigations of the archetypal lower respiratory tract infection caused by this organism, known as lobar pneumonia.

This was a condition whose prognosis was poor, onset generally followed (and indeed follows still) an episode of lesser illness, often viral. As the condition proceeds, rapid inflammation of the tissues of the lung takes place, followed by influx of fluids into the major lobes of the lung (or lobar consolidation), an extremely painful experience for the victim. The next stages of the disease include 'red hepatisation' where the lung tissues appear red in colour due to the influx of red blood cells from the distended vascular tissues. This is followed by 'grey hepatisation' caused by accumulation of fibrin associated with red and white blood cell degradation. Further outward symptoms vary depending on the severity of the illness, but include shaking chills, chattering teeth, and production of rust-coloured or greenish mucus. These symptoms are associated with high temperature, profuse sweating and mental perturbation. The blood may also become depleted of oxygen, leading to discolouration of the lips and nail beds.

Death can occur in a number of ways; anoxia brought about by compromised gaseous exchange at the respiratory mucosa, complications arising from secondary infections by other organisms, or spread of the pneumococcus to other sites via the circulatory system. This latter route can subsequently lead to pneumococcal meningitis and other invasive conditions.

Investigation into the causes of this disease were initiated most notably by Jurgensen in 1874 when he established the link between this condition and other diseases caused by infectious agents. He noted that "The assumption of a specific aetiologic agent is necessary, croupous (literally, coughing) pneumonia belongs, then, to the group of infectious diseases..." (Foster, 1970). This prophecy was borne out after the introduction of Robert Koch's plate technique for the isolation of micro-organisms in pure culture, when a cascade of information regarding the agents of disease came about.
There is some variation in the early literature regarding exactly who was the first to isolate and identify the pneumococcus. In1886 Albert Fraenkel, a disciple of Koch's was credited for growing the pneumococcus in pure culture and noting its relationship to disease (Bulloch, 1960, Brock, 1988), however, the organism was almost certainly described before this. In 1882 C. Friedlander (working with the 31 year old Dane, H.C.J. Gram) described diplococci in lung sections taken post mortem. The same year Leyden

and Gunther reiterated this observation, also noting that the cocci were capsulated (this time isolated from a living subject).

At around this time the work of Friedlander became confused by the accidental inclusion of what is now known to be *Klebsiella pneumoniae* in his experimental samples. These capsulate organisms are found in the lung, grow in pairs, and crucially, more successfully than pneumococci on the medium used by Friedlander.

The final chapter of this early work in *S. pneumoniae* was written by A. Weichselbaum, who extended the appreciation of the pathogenic ability of the organism when he was able to implicate the pneumococcus in cases of meningitis, endo- and pericarditis and arthritis.

1.2 - Pneumococcal infection in the twentieth century

The pneumococcus still presents us with very high levels of morbidity and mortality today. The UK Public Health Laboratory Service (PHLS) cites annual rates of invasive (meningitis and bacteraemic) pneumococcal infection at around 4000 to 5000 cases per year (data based on that available from the PHLS website at http://www.phls.co.uk/facts/streppn.html). Data from the CDC in the USA cites half a million pneumonia cases, 3000 meningitis cases, 50,000 bacteremia cases, 100,000-135,000 hospitalisations for pneumonia and 7 million otitis media cases annually (http://www.cdc.gov/ncidod/dbmd/diseaseinfo/streppneum_t.html).

The ability of this pathogen to remain such an important agent of disease (despite advances in the comprehension of the molecular basis of the interplay between host and pathogen) rests on two key elements of the biology of the organism.
The production of virulence determinants in the form of the capsule, and protein virulence determinants enables the organism to evade host immune responses and bind to and invade the host tissues (Watson and Musher, 1990, Lee *et al.*, 1991, Frost *et al.*, 1996, Gray, 2000). This is complemented by the naturally transformable nature of the organism. The ability to take up exogenous DNA, and incorporate it into the genome affords the pneumococcus a high degree of genetic, and subsequently, antigenic

plasticity. This also renders the organism responsive to selective pressure, be it in the form of the attentions of the host immune system or antimicrobial chemotherapy.

1.3 – Genetic plasticity in the Pneumococcus - Transformation

In 1928 Griffith established the link between pathogenicity and the capsule by noting that non-encapsulated (or 'rough') strains were avirulent in an intravenous mouse model of infection. He later conducted the classical experiment in which he co-inoculated mice with two different strains; a live rough strain derived from a serotype 2 strain which had lost its capsule, and a smooth (capsular) serotype 3 strain which had been heat killed. The result was that some of the mice succumbed to infection, and the strains isolated from the blood of these mice were characterised as serotype 3 (Griffith, 1928). Hence he had implicated the transfer of the capsular phenotype of the dead type 3 strain to the live genotype 2 strain. 16 years later (in the experiments of Avery, MacLeod and McCarty) DNA was identified as the substance mediating the transfer of this phenotype (Avery *et al.*, 1944).

1.3.1 – Regulation of competence for genetic transformation

Today there is a good deal understood about the molecular basis for genetic variation by transformation and recombination in the pneumococcus. In brief, the events surrounding this process involves binding exogenous dsDNA at the cell surface, degradation of one strand of the incoming molecule, followed by translocation of the ssDNA into the cytoplasm in the 5' to 3' direction. The translocated strand is methylated by DpnA and coated with Single Strand Binding protein (Ssb) (Morrison, 1978, Morrison and Marinelli, 1979, Lacks *et al.*, 2000). These steps are followed by RecA mediated recombination into the genome. Key elements of these processes are described in more detail below and illustrated in figure 1.1.

The significance of natural transformation is illustrated in the context of target mediated resistance to the first described class of antimicrobial agents, the penicillins. Resistance to the penicillins has arisen by alterations to the genes encoding the Penicillin

Binding Proteins (PBPs), the molecular targets for this class of drugs. The initial events surrounding the emergence of resistant clones seems to have involved development of low affinity forms of PBPs in related oral streptococci (presumably under chronic selective conditions over time), followed by horizontal gene transfer of these loci into pneumococci (Dowson *et al.*, 1993, Coffey *et al.*, 1993).

The nature of the selective pressure provided by antibiotic use and the opportunities for genetic transfer between related streptococci has rendered the pneumococcus frequently resistant to penicillin. Clonal groups of penicillin non-susceptible pneumococci have come to characterise the global distribution and impact of pneumococcal disease (described in more detail in section 1.4).

The onset of competence is tightly regulated in the growing population of pneumococci, and represents a classical two component regulatory system, with added layers of regulatory complexity which are under active investigation by a number of groups.

Pneumococci become competent during early log phase and only remain competent for around 40 minutes (Morrison, 1997, Ween and Haverstein, 1999), and competence was hypothesised to be under the control of an 'activator' substance (Tomasz, 1965).

This substance was later characterised as an oligopeptide produced and secreted by pneumococci during growth, and named CSP (for Competence Stimulating Peptide). The gene encoding this peptide, termed *comC* was isolated using reverse genetics starting with the peptide sequence (Haverstein and Morrison, 1995).

comD and *comE* were found to encode the 2 component regulator/response apparatus of this system (Pestova *et al.*, 1996, Cheng *et al.*, 1997).

Subsequent studies have shown that there is more than one class of pheromone produced, complemented by structural modifications of the ComD protein associated with a given pherotype (Claverys *et al.*, 2000).

The *comA* and *comB* genes encode the CSP export proteins (Morisson, 1997). The ComA protein is a member of the ATP dependent transport proteins, and the *comAB* promoter is a target for the phosphorylated ComE response regulator (Cheng *et al.*, 1997). When exogenous CSP reaches a threshold level in the growing culture, genes

required for its own expression, export and detection are all up regulated (Lacks *et al.*, 2000).

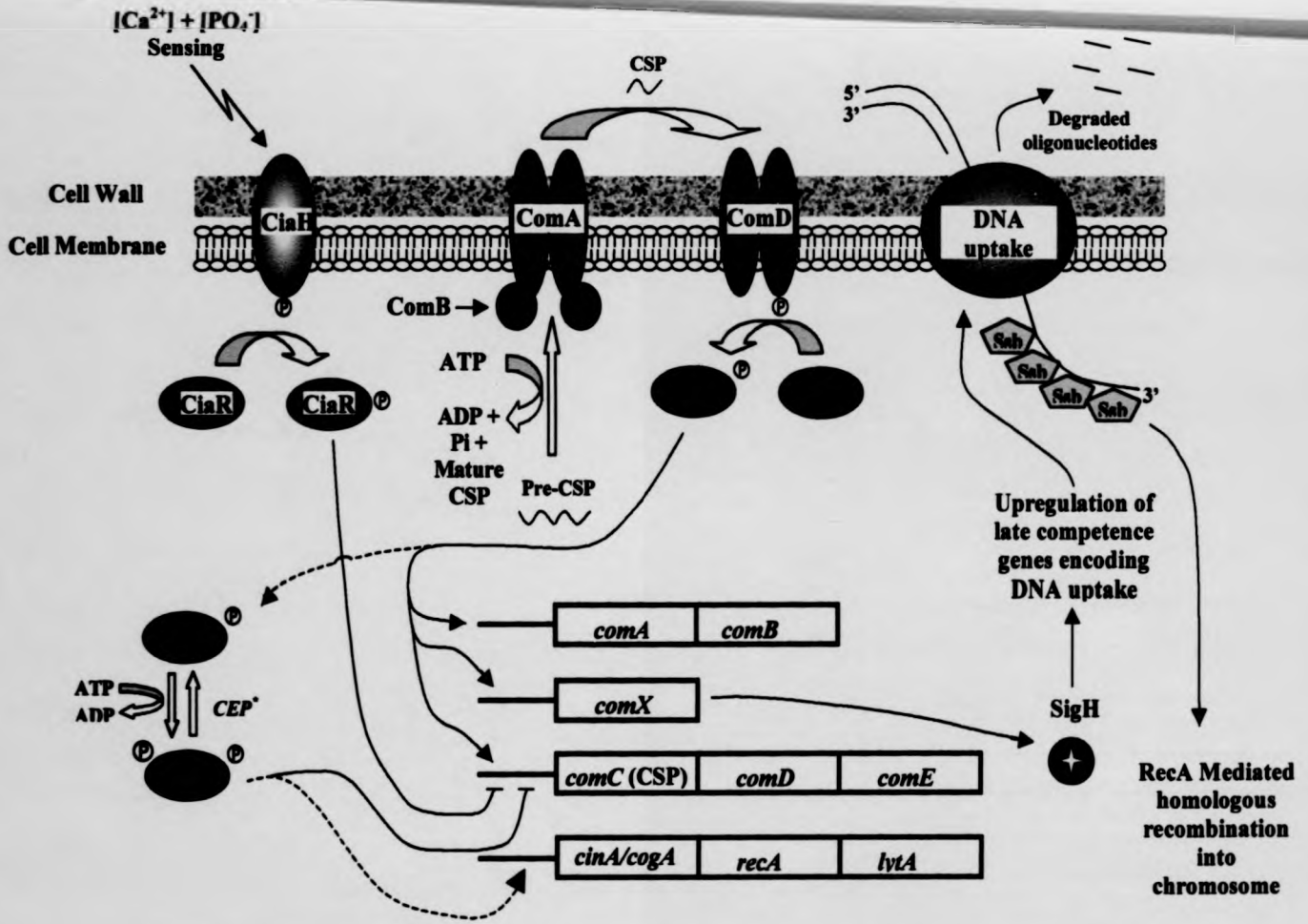


Figure 1.1 - Summarising key elements of the natural transformation process. Adapted from Lacks *et al.* (Lacks *et al.*, 2000). See text for details. *CEP: hypothetical ComE specific Phosphatase (Alloing *et al.*, 1998).

Multiple regulatory systems have been found that demonstrate specific metabolite requirements before competence for genetic transformation can develop. These include a requirement for O_2, sensed via NADH oxidase (Auzat *et al.*, 1999), cations such as Ca^{2+}, Zn^{2+} and Mn^{2+} (Fox and Hotchkiss 1957, Dintilhac *et al.*,1997) and phosphate (Novak *et al.*, 1999). The *ciaH/R* two component system is thought to be involved in regulation of competence in response to calcium and phosphate levels, repressing the early competence genes when these ions are at sub-optimal concentrations (Giammarinaro *et al.*, 1999).

Another key element in the competence regulatory system is the SigH protein, an alternative sigma factor encoded by the *comX* gene. SigH and can be considered as the regulatory link between the early competence genes and those involved in the later stages of the process (Lee and Morrison, 1999). SigH functions by recognising and binding to the consensus sequence 5'-TACGAATA-3' at position –10 with respect to the transcriptional start point. This sequence is termed the 'cin-box', and precedes genes encoding components of the DNA translocation complex, Ssb and the nuclease responsible for degrading the incoming 5' to 3' strand of DNA (EndA) (Lacks *et al.*, 2000).

A recent study by Rimini *et al.* has investigated the temporal distribution of gene expression during competence induction using cDNA hybridisation to high-density genomic microarrays. This approach enabled the authors to identify genes that are up regulated and down regulated during the development and shutdown of competence. Up regulated loci include the known competence genes (described above) plus others implicated in diverse processes including DNA modification and central metabolism. Down regulated genes notably include ribosomal proteins, indicating a transient shutdown of protein synthesis during competence (Rimini *et al.*, 2000). Whilst of great interest, a further discussion of the implications of this work are beyond the scope of this chapter.

1.3.2 – DNA uptake and homologous recombination

Unlike other naturally transformable bacteria, such as *Haemophilus influenzae*, *S. pneumoniae* is promiscuous in that it does not require a species-specific sequence to identify DNA suitable for uptake, thus opening up the possibility of inter-species genetic exchange.

As the incoming strand enters the cell, it undergoes methylation at 5'-GATC-3' motifs by the product of the *dpnA* gene, one of the late competence genes expressed under the control of SigH (Lacks *et al.*, 2000). Claverys *et al.* suggest that this methylation is required to protect the inserted, single stranded donor DNA from restriction by *Dpn*II prior to resolution and replication of the double stranded recombinant DNA (Lefrancois *et al.*, 1998, Claverys *et al.*, 2000).

Integration into the host chromosome has been shown to be dependent upon RecA (MortierBarriere *et al.*, 1998), which is transported to the inner face of the cell membrane by the product of the *cinA* gene (Masure *et al.*, 1998).

The proposed mechanism of recombination is complex and involves a number of proteins, of which RecA is thought to play a principal role (Camerini-Otero and Hsieh, 1995). A mismatch repair (MMR) system operates in *S. pneumoniae* in the form of HexAB (the *hexAB* genes in the pneumococcus are homologous to the *mutSL* genes of *E.coli*). While this system efficiently recognises and repairs single base mismatches, it is prone to saturation in the presence of large numbers of mismatches (Humbert *et al.*, 1995). So, when errors occur during normal chromosomal replication, they are efficiently repaired by Hex. However, when the cell takes up DNA during transformation, and the recombination process is underway, Hex does not present a significant barrier to integration and the establishment of novel sequences. This may represent an adaptation of the pneumococcus to its recombinogenic lifestyle, and the genetic plasticity that it affords (Majewski *et al.*, 2000).

It is estimated that competent pneumococci can take up several DNA fragments per cell, accounting for up to 5% of the pnemococcal genome. Estimates suggest that the

probability of recombination of such fragments into the genome is 0.1-0.5 for 2-20% divergent DNA, rising to 1 for homologous DNA (Claverys *et al.*, 2000).

Observations from our laboratory and others indicate that inter-species recombination has been involved in the evolution of several pneumococcal genes (S. King – Unpublished data, Poulsen *et al.*, 1998, Whatmore and Dowson 1999). This has implications for vaccine design, as protein based formulations will conceivably select for escape variants unless suitable precautions are taken. Current attentions are centred on searching for a stable pneumococcal protein for inclusion in the next generation of pneumococcal vaccines.

1.4 - Incidence of *S. pneumoniae*: Clones and serotype prevalence

Despite the pathogenic capacity of the pneumococcus, it is able to persist in the host population unnoticed. The organism can be carried in the nasopharynx of carriers in an asymptomatic state in many cases. It is often thought of as part of the normal flora of the human upper respiratory tract. Carriage rates are estimated to be 5-10% of a healthy adult population and 40-50% of infants under the age of 24 months (Gray, 2000), however these figures are subject to seasonal and geographic variation.

A key determinant of host susceptibility to pneumococcal infection is the immune status of the host. A minor episode of viral infection is often enough to enable the pneumococcus to take hold, indeed the activity of influenza virus neuraminidase has been implicated in the unveiling of host cell receptors for pneumococcal adhesion (Plotkowski *et al.*, 1986, AlonsoDeVelasco *et al.*, 1995). Alternatively, the sub-optimal immune responses of young children or elderly individuals may provide the pneumococcus with the opportunity to invade. Other paths to immunocompromisation include infection with HIV, impaired function or loss of the spleen, and suppression of the immune response by chemotherapeutic drug use. In this light, the pneumococcus can be regarded as an opportunistic pathogen.

In all, 90 serotypes are known to exist, some of which are particularly associated with invasive disease. Certain capsular polysaccharides are associated with invasive infections and delineated on the basis of the target population. For example, serotypes 1, 3, 4, 7, 8

and 12 are commonly isolated from invasive disease in adults, whereas serogroups 6, 14, 19 and 23 are more common in young children (Gray, 2000).

A key feature of pneumococcal incidence is the occurrence of clonal population growth. The pneumococcus is able to undergo the usual processes of genetic drift by point mutation of its genes. It is also able to undergo genetic shift via horizontal transfer of genetic material with other bacteria.

Pneumococcal progeny which have acquired a particular combination of genetic elements (which gives them a selective advantage over their competitors) are able to proliferate rapidly, leading to the appearance of pneumococcal clones. The development of antibiotic resistance (notably to penicillin) has driven the appearance of such clones.

Techniques used to describe clonal relatedness include antibiogram definition, chromosomal techniques such as pulsed-field gel electrophoresis and arbitrarily primed PCR, more sensitive techniques include Restriction Fragment Length Polymorphism (RFLP) analysis of particular genes, and more recently, Multi-Locus Sequence Typing (MLST) (Enright and Spratt, 1998).

Much of the classical work on the spread of antibiotic resistant clones originated from the study of the Spanish serotype 23F, 6B clones and the 9V clone thought to have originated in France (Munoz *et al.*, 1991, Enright *et al.*, 1999). Subsequent studies revealed the existence of another three multiply antibiotic resistant clones of serotypes 14, 15F and 19F, also from Spain (Coffey *et al.*, 1996). Molecular approaches have traced the spread of some of these clonal strains across the globe. Derivatives of the Spanish 23F clone (Spain^{23F}-1) is known to have spread to the USA (Munoz *et al.*, 1991), Hong Kong (Ip *et al.*, 1999), continental South America (Castaneda *et al.*, 1998), Mexico (EchanizAviles *et al.*, 1998) and Korea (McGee *et al.*, 1997). This clone also appears to have spread to Belgium, via France, undergoing some variation at the genome level as it has done so (Ferroni *et al.*, 1996, HoefnagelsSchuermans *et al.*, 1999). These observations have secured the pandemic status of this clone.

Much emphasis is placed on the identification of a particular clone on the basis of the serotype of the members of that clone. The genetic (and hence phenotypic) plasticity of the pneumococcus has been hinted at in the case of the evolution of the PBPs. It is not surprising then, that the nature of the capsule can alter by horizontal gene transfer also, a

phenomenon known as Capsular Switching. The molecular techniques employed to identify the early emerging clones have also helped reveal instances where the genes encoding the capsule (and hence serotype, see below) have rearranged with those of other pneumococci. Certain instances have been studied in molecular detail (Coffey *et al.*, 1998, Coffey *et al.*, 1999, Ramirez and Tomasz, 1999a) and capsular switching is thought to be a common mechanism of phenotypic variation.

While the ability of successful pneumococcal clones enable the organism to evade certain forms of antimicrobial therapy, initiation of invasive disease and the evasion of the attentions of the host immune system are conferred by other attributes of the pneumococcus.

1.5 - Virulence determinants of *S. pneumoniae*

Before embarking on a description of the pneumococcal virulence determinants, it may be beneficial to briefly describe the architecture of the pneumococcal cell.

The cell membrane is surrounded by the cell wall responsible for the Gram positive phenotype. This is in turn overlaid by the capsule in those cells expressing capsular polysaccharide (see below).

The cell wall consists of crosslinked peptidoglycan layers (GlcNac-β1-4-MurNac) linked to its cognate stem peptide (Gray, 2000). Imbedded in this matrix is the cell wall polysaccharide (CWPS) or C polysaccharide. This is a complex teichoic acid containing phosphorylcholine residues (AlonsoDeVelasco *et al.*, 1995).

The breakdown products of the peptidoglycan complex are highly inflammatory (usually released after autolysin mediated autolysis – see section 1.5.3.1), and are thought to be responsible for many of the toxic effects of pneumococcal infection (Gray, 2000). The cell wall is also the location of many pneumococcal proteins such as the PBP's and some of the protein virulence determinants to be described in detail below.

1.5.1 – The capsule

The capsule is sometimes considered to be the primary virulence determinant of the pneumococcus. It contributes to disease by inhibiting clearance by host phagocytes. This is reflected in the 10^5 fold difference in LD_{50} between encapsulated and acapsular strains in a murine challenge model (Watson and Musher, 1990).

The capsule is present overlaying the cell wall. Interspersed with its polysaccharide chains are proteins such as the pneumococcal surface proteins (Psp's) and cell wall constituents such as teichoic acid (TA or C-polysaccharide), lipoteichoic acid (LTA) and phosphorylcholine. Most of these elements are potentially immunogenic. Antibody responses to the capsular polysaccharide are detected in immunocompetent individuals during the course of infection, however, such T-cell independent responses are poor in two key populations, namely the elderly, and the very young (less than 24 months).

The capsule is not ubiquitous and under some circumstances is lost. The amount of capsule produced can vary at any one time, and it has been suggested that an isolate of *S. pneumoniae* should be considered to contain a mixture of capsule phenotypes (Weiser and Kapoor, 1999). The presence or absence of the capsule bestows different phenotypes upon the organism in terms of its relationship with the host. Capsulate pneumococci are resistant to phagocytosis, whereas acapsular cells are not. However, the presence of the capsule presents a barrier to cytoadherence with in the nasopharynx by the occlusion of structures on the pneumococcal cell surface responsible for binding to host macromolecules. Hence the balance between persistence in the host and clearance by immune surveillance, and that between colonisation of the host mucosa and loss by mass action from the respiratory tract is determined by the processes which regulate capsule production within the pneumococcus.

The biosynthetic genes required for capsule synthesis are closely linked and are arranged as operons. The caspsular biosynthetic gene clusters of strains of different capsular serotypes are located in the same chromosomal locations (Claverys *et al.*, 2000), and there has been much work investigating the nature of the capsular biosynthetic gene clusters. To date the loci encoding production of the type 23F (Morona *et al.*, 1999b)

Ramirez and Tomasz, 1999b), 2 (Iannelli *et al.*, 1999), 8 (Munoz *et al.*, 1999), 19 (Morona *et al.*, 1999a), 37 (Llull *et al.*, 1999), 3 (Arrecubieta *et al.*, 1995, Arrecubieta *et al.*, 1996), 1 (Munoz *et al.*, 1997), 14 (Kolkman *et al.*, 1997) and 33F (Llull *et al.*, 1998) have been described. The complexity of individual capsular loci correlates approximately with the chemical complexity of the polysaccharide produced.

1.5.2 – Immune responses to pneumococcal antigens

Host clearance of pneumococci is thought to depend upon effective opsonisation (both by antibodies and complement components), followed by phagocytocis and subsequent killing in the phagolysosomal compartment (AlonsoDeVelasco *et al.*, 1995, Tu *et al.*, 1999).

Exogenous antigens capable of eliciting a host antibody response can be broadly divided into two categories; the thymus dependent (TD) antigens, and the thymus independent (TI) antigens. The nature of these antigens and that of the host response to them differ somewhat. The TD antigens are often proteinaceous, and the host response to them is anamnestic.

The TI antigens are generally high molecular weight macromolecules, which are poorly metabolised in the host. This class of antigens includes polysaccharides, polypeptides and polynucleotides. Responses to these antigens are not anamnestic, and importantly, are age dependent.

The available pneumococcal vaccines are multivalent polysaccharide preparations, which constitute TI antigens. While the host response to TI antigens in the healthy adult is effective, this is not the case in other groups of individuals, making them particularly vulnerable to pneumococcal disease. Asplenic patients, those who suffer frequent RTI's (respiratory tract infections), HIV patients, the elderly, and the young comprise these high-risk groups. Geriatric production of anti-polysaccharide (anti-PS) ab is seen to decline in individuals from the age of 50-60 onwards. Paediatric production of anti-PS ab is only observed after the age of 4-5 years (as seen with the prevalent serotypes affecting children, 6A, 14, 19F and 23F) (Douglas *et al.*, 1983). Thus it is apparent that the available polysaccharide vaccines are not particularly useful to these groups of

individuals. The current drive of vaccine research includes an investigation of the efficacy and antigenic stability of pneumococcal proteins for inclusion in the next generation of pneumococcal vaccines, consisting of PS antigens conjugated to protein carriers.

1.5.3 – Protein virulence determinants

The work presented in this thesis is based around a specific field of research into pneumococcal protein virulence determinants (PVD's). These afford the pneumococcus the ability to invade the tissues of the host and cause disease.

1.5.3.1 – The Choline binding proteins and Autolysin

The pneumococcal cell wall, as mentioned above, is comprised of a complex mixture of proteins, peptides and teichoic acids (Navarre and Schneewind, 1999). Many of the proteins found immobilised the pneumococcal cell wall can be broadly categorised as those which bear the classical gram positive cell wall anchor motif (LPXTG), and those which bear choline binding domains. Pneumococcal teichoic acids are rich in choline, however the pneumococcus is unable to synthesise this compound, for which it has an absolute requirement (Yother *et al.*, 1998).

To date a number of pneumococcal choline binding proteins (Cbp's) have been described. These include the major autolysin, LytA (Garcia *et al.*, 1986, Lopez *et al.*, 1986, Whatmore and Dowson 1999), pneumococcal surface protein A (PspA) (Talkington *et al.*, 1991), a murein hydrolase, LytB (Garcia *et al.*, 1988), a lysozyme, LytC (Garcia *et al.*, 1999), PcpA, which is thought to play a role in pneumococcal adhesion (Sanchez-Beato *et al.*, 1998) and an adhesin, CbpA (Rosenow *et al.*, 1997) also termed SpsA or PspC (Brooks-Walter *et al.*, 1999).

Mutagenesis studies have shown that some of the Cbp's behave as adhesins. These include CbpA, which is thought to interact with GalNAc-β1,4–Gal and GalNAc-β1,3–Gal receptors on host cells.

The precise function of PspA is not known. Recent studies have implicated PspA in blocking the activation of the alternative complement pathway at the stage of Factor B mediated activation of C3b. Tu *et al.* have shown that loss of PspA leads to rapid clearance of pneumococci from infected mice (Tu *et al.*, 1999).

The first Cbp to be characterised was the major autolysin, LytA. This protein is an N-acetyl-muramoyl-L-alanine amidase, and is involved in degradation of peptidoglycan. Like many of the Cbp's, it has a modular organisation, with the (now well characterised) choline binding domain at the C-terminus of the molecule and the catalytic portion at the N-terminus.

The nature of the contribution of LytA to virulence is thought to vary depending upon the route of inoculation in animal models, reflecting different disease states in the human host (Berry *et al.*, 1989, Sato *et al.*, 1996, Whatmore and Dowson, 1999).

LytA is known to contribute to virulence in two ways, firstly by promoting inflammation by the release of cell wall breakdown products. Secondly, the key pneumococcal cytotoxin, pneumolysin (see section 1.5.3.2, below) is not actively exported, rather it accumulates in the pneumococcal cytoplasm and is released by autolysin mediated lysis of the cell.

1.5.3.2 – Pneumolysin

Pneumolysin is one of the classical pneumococcal PVD's, and has been the focus of a considerable amount of investigation. The pneumolysin monomer is a 53kDa haemolysin that targets and inserts into cholesterol-containing host cell membranes. Here the monomers polymerise, forming a transmembrane pore, leading to host cell lysis (Paton *et al.*, 1993).

Pneumolysin belongs to the family of thiol-activated toxins, and is released from the pneumococcal cytoplasm following general cell lysis mediated by autolysin. It is the action of this class of enzymes that is responsible for the haemolytic phenotype seen when pneumococci are grown on blood agar media.

At lower concentrations, pneumolysin is known to have numerous effects on the host, including impeding epithelial cilliary function (Steinfort *et al.*, 1989) and enhanced

inflammation due to the release of tumour necrosis factor-α and interleukin-1β from host monocytes (Houldsworth *et al.*, 1994). Neutrophil migration and activity is affected as well as lymphocyte replication and antibody synthesis (Paton and Ferrante, 1983, Ferrante *et al.*, 1984).

Pneumolysin also interacts with soluble components of the immune system, resulting in inappropriate activation of complement, leading to further inflammation. (Mitchell *et al.*, 1991, AlonsoDeVelasco *et al.*, 1995, Benton *et al.*, 1997). This is thought to be a consequence of the structural similarity of domain 4 of the pneumolysin monomer to the Fc portion of IgG (Rossjohn *et al.*, 1998).

The clinical features of pneumococcal pneumonia can be induced by the addition of purified pneumolysin to the rat lung (Mitchell *et al.*, 1997). Similarly, when isogenic pneumolysin negative strains are used in a mouse lung infection model, wild type growth can be restored by the addition of exogenous pneumolysin. A similar effect was seen in the case of irreversible hearing loss after episodes of pneumococcal meningitis (Winter *et al.*, 1997).

As a noteworthy aside, it is perhaps important to mention the finding that there appears to be more than one pneumococcal haemolysin. The locus encoding pneumolysin, *ply*, appears to be redundant to loci encoding at least two further haemolysins. The contribution of these alternative haemolysins to virulence is currently under investigation in our laboratory (S. Colby and D. Sturgeon – unpublished data).

1.5.3.3 – Neuraminidases

Neuraminidases are sialidases that cleave N-acetylneuraminic acid (NeuNAc) from glycoproteins, mucin and gangliosides. The physiological consequences of such an activity during pathogenesis involves reduction of mucus viscosity and the unveiling of potential receptors for bacterial attachment to host epithelia (Mitchell *et al.*, 1997, Tong *et al.*, 2000).

S. pneumoniae is known to have at least two genes encoding neuraminidases, designated *nanA* and *nanB* (Berry *et al.*, 1988, Camara *et al.*, 1991, Camara *et al.*, 1994).

The first description of a pneumococcal neuraminidase involved screening a cosmid library for clones which were able to encode proteins able to degrade the fluorogenic neuraminic acid analogue, 2'-(4-methylumbeliferyl)-α-D-*N*-acetylneuraminic acid (MUAN).

NanA and NanB differ in terms of their pH optima, and the two proteins possess different sorting sequences. NanA has an N-terminal cleavable signal sequence, and a C-terminal cell wall anchor motif, whereas NanB has a cleavable N-terminal signal sequence but no cell wall anchor at its C-terminus. This suggests the differential use of cell associated and secreted neuraminidases with different enzymatic properties.

The contribution of neuraminidase to virulence has been assessed in some animal models by insertion-duplication mutagenesis at the *nanA* locus (Winter *et al.*, 1997). This mutant, named ΔNA1 was shown to be deficient in nasopharyngeal colonisation and induction of otitis media in the chinchilla model (Tong *et al.*, 2000). A similar attenuation of virulence was observed in a murine pneumonia model (Mitchell *et al.*, 1997), but not the guinea pig meningitis model (Winter *et al.*, 1997). A role for other neuraminidase(s) in these systems was ruled out on the basis of negligible degradation of MUAN by the ΔNA1 strain.

1.5.3.4 – IgA1 Protease

The IgA1 proteases contribute to virulence by interfering with the opsonisation process of mucosal IgA1, the predominant immunoglobulin of the upper respiratory tract. By cleaving the dimeric molecule at the hinge region, the IgA1 proteases uncouple antigenic recognition from the effector responses of the immune system.

Such proteases represent a diverse group of enzymes, found in a wide variety of organisms including *H. influenzae, Neisseria meningitidis, N. gonorrhoeae, Streptococcus sanguis, S. mitis, S. oralis* and *S. pneumoniae*.

The IgA1 protease from the pneumococcus is thought to act in concert with neuraminidase, with the latter enzyme disrupting the glycosylated regions of the IgA1 molecule, making it more susceptible to proteolysis by IgA1 protease (Reinholdt and Kilian, 1996).

The structure of the pneumococcal *iga* gene contains some interesting features; a cleavable signal sequence is evident at the N-terminus (Poulsen *et al.*, 1996, Wani *et al.*, 1996). Downstream from this there are a number of tandem repeats. The nature and extent of reiteration of these repeats is variable among the IgA1 proteases. The C-terminal portion of the gene, thought to contain the catalytic centre, harbours the zinc-binding motif, $HEMTH(X)_{20}E$ (Poulsen *et al.*, 1996, Poulsen *et al.*, 1998).

The novel location of a cell wall anchor motif at the N-terminal region of the protein is an interesting feature, and the distribution of hydrophobic regions around the anchor motif, and the position of a proline-rich region (often associated with surface proteins) suggests an alternative topology of the protein at the cell surface, with three transmembrane spanning regions (Poulsen *et al.*, 1998).

Sequence analysis revealed that the streptococcal *iga* loci display mosaic structures, and are evolving by transformation-mediated homologous recombination between the species *S. oralis, S. mitis* and *S. pneumoniae*, but not *S. sanguis*.

A phenotypic analysis of IgA1 protease activity in a wide range of strains was undertaken by Reinholdt and Kilian (Reinholdt and Kilian, 1997). They found that activity of this enzyme varied widely throughout the streptococci tested, with no specific association of IgA1 protease activity, with either commensal or pathogenic organisms.

1.6 – *in vivo* studies of pneumococcal protein virulence determinants

The purpose of studying PVD's *in vivo* is twofold; firstly to assess the contribution of individual virulence determinants to pathogenesis in different modes of infection. The second interest is evaluation of pneumococcal proteins as immunogens. The currently available vaccines belong to the Pneumovax class. These compositions consist of polyvalent arrays of capsular polysaccharides. As mentioned in section 1.5.2, such vaccines are only efficacious in immunocompetent individuals. Their applicability in other populations is restricted by the inability of those individuals to mount an appropriate and long-term immune response to polysaccharide antigens.

To take these issues in turn, the individual contribution of the principal virulence determinants in particular model systems has been mentioned in the relevant sections above. It is perhaps noteworthy to mention the *in-vivo* approach to identification of

virulence determinants taken by Polissi *et al.* (Polissi *et al.*, 1998). This approach involved a powerful signature tagged mutagenesis (STM) technique to isolate mutants with attenuated virulence. A population of mutagenised pneumococci was generated by recombination after suicide vectors were introduced by transformation, containing random pneumococcal genome fragments, plus a unique oligomeric 'tag'.

Virulence attenuated mutants from either pneumonia or septicaemia models were identified after Southern based screening of the bacteria recovered from the mouse lung or bloodstream respectively against individual mutants generated at the start of the experiment (the regions containing the unique tags were used in the hybridisation steps). Mutants which were present in the *in-vitro* pools, but not the *in-vivo* pools were characterised by Inverse PCR (iPCR – see Chapter 2, section 2.2.1 for a description of the technique).

The genes identified by this approach were numerous, and some differences were noted between those genes isolated from the pneumonia and septicaemia models. The genes identified after the first round (intranasally inoculated) of selection were grouped into 4 classes. The first included the known virulence determinants hyaluronate lyase, neuraminidase, autolysin and IgA1 protease. Interestingly, pneumolysin was not identified as a requirement in this screen, unlike other pneumonia models using pneumolysin negative mutants (Mitchell *et al.*, 1997). The authors suggest that this may be a consequence of trans complementation by wild type expression of Ply from other strains in the pool. This would still reflect a requirement for LytA (which remains cell associated), as was the case in this experiment. The inclusion of IgA1 protease in this group is also interesting. It is understood that this enzyme has no activity against murine IgA1 (Proctor and Manning 1990), therefore IgA1 protease must play an alternative role in this system.

The second class of genes identified as required were those involved in various metabolic functions. These included genes involved in purine biosynthesis, glutamine biosynthesis, phosphate fixation and anaerobic metabolism. This suggests that pneumococcal genes involved in survival under nutrient limiting conditions are required in this model. The final 2 classes of genes included proteases, transport systems and genes involved in DNA recombination and repair.

When these mutants were tested in the septicaemia model it was found that there was no longer a specific requirement for hyaluronate lyase, neuraminidase or autolysin. Therefore it appears that these enzymes play a role in colonisation and persistence at the mouse respiratory epithelial mucosa, but are not required after the transition across the lung-blood barrier, as reflected in the mouse septicaemia model. The strengths of this experimental approach lie in the inclusion of the multitude of factors present in the *in-vivo* selection system, which are not present under *in-vitro* culture conditions. Thus the interplay between host and pathogen is represented in this model.

Berry and Paton examined double mutants of *nanA, hyl, pspA, lytA* and *cbpA* in a pneumolysin negative background (Berry and Paton, 2000). They found that in the mouse septicaemia model, the pneumolysin knockout (ΔPly), was significantly less virulent than the wild type strain (D39 serotype 3 isolate), whereas Hyl^- and $NanA^-$ single knockouts were not. The $PspA^-$ mutation resulted in moderate attenuation of virulence. Interestingly, when the double mutants ΔPly-$Hyl^-$and ΔPly-$PspA^-$ were tested, they found that these strains were significantly less virulent than the ΔPly strain alone. This observation leads to the interpretation that the combined influence of pneumococcal PVD's constitutes a system where the whole is greater than the sum of the parts.

The second central aim of *in-vivo* studies is to assess the immunogenicity of pneumococcal protein antigens, and their potential efficacy for inclusion in the next generation of conjugate vaccine preparations that include both proteinaceous and polysaccharide components.

As mentioned in section 1.5.2, individuals who are at high risk from pneumococcal infection do not possess the ability to respond well to TI antigens, including polysaccharides. This renders the Pneumovax class of vaccines ineffective in these groups. In order to address this problem, the phenomenon of enhanced recognition of PS antigens in the context of a protein carrier molecule has been exploited. This approach has been highly successful in protection against *H. influenzae* type b (Mitchell *et al.*, 1997), where anamnestic responses to polysaccharide antigens has been achieved.

Primary work in this field involved conjugation of streptococcal polysaccharides to the heterologous diptheria and tetanus toxoids. These were selected on the basis of their Th cell activating ability, and were effective in animal models immunised with GBS type

3 and serotype type 4 pneumococci (Peeters *et al.*, 1991, Paoletti *et al.*, 1992). Drawbacks associated with this approach include sequestration of the antibody response by excessive anti-carrier antibody production. This has been shown to result in suppression of immune responses to subsequent formulations using the same carrier in mice (AlonsoDeVelasco *et al.*, 1995).

In this light it can be seen that pneumococcal carrier proteins would provide attractive alternatives to non-pneumococcal carriers. If carrier proteins or peptides can be identified which confer appropriate T-helper and B cell stimulation, and confer antibody mediated responses to the hapten polysaccharide moieties, the potential advantages are twofold; early-onset antibody mediated opsonisation against PS epitopes, followed by protein mediated opsonisation, and/or abrogation of function of pneumococcal toxins.

An obvious limitation to the use of conjugation of PS to carrier peptides or proteins is the number of haptens attached per preparation. This could be simply addressed using a knowledge of serotype prevalence within the given target population (segregated on the basis of geography or target group).

Another potential drawback is the emergence of carrier-escape variants within the population. Where the TD anti-carrier ab response is able to exert selective pressure on the population of pneumococci as a whole, transformation/homologous recombination mediated antigenic escape becomes a possibility.

To this end, knowledge of the population diversity of carrier candidates is required. Pneumococcal genes including *lytA*, *pspA* and *nanA* are all thought to have evolved by homologous recombination between divergent alleles (Swaitlo *et al.*, 1997, Garcia *et al.*, 1999, Whatmore and Dowson, 1999 and S. King – unpublished data). In this study the diversity at the *hyl* locus is examined to establish the extent and nature of variation in order to examine the potential use of the protein (or derivatives of it) in new conjugate pneumococcal vaccines.

1.7 – Hyaluronate lyase

The remainder of this introductory chapter will focus on the pneumococcal hyaluronidase, more accurately termed hyaluronate lyase. This will begin with an

introduction to the substrate, followed by a discussion of the class of enzymes responsible for its degradation. This will be followed by a description of the pneumococcal enzyme. An outline of the aims and perspectives of the work presented here will follow an introduction to the available assays for hyaluronic acid degradation.

1.7.1 – Occurrence of hyaluronic acid in the host

Hyaluronic acid (HA) is a member of the glycosaminoglycan (GAG) family of carbohydrates. This class of polysaccharide is characterised by disaccharide repeat units, one component of which is always an amino sugar. In many cases the amino sugar is sulphated at the C4 position, although this is not the case for HA. The remaining component of the disaccharide is an uronic acid. The GAG's fall into 3 distinct classes, characterised by their disaccharide composition and linkage pattern. The polymaltose group contains heparan sulphate, the polylactose group includes the chondroitin and keratan sulphates and the cellobiose class includes HA.

The GAG's are often associated with a core protein, forming the proteoglycan component of the extracellular matrix. The other major components of the extracellular matrix are proteins such as collagen and elastin. Together these major components form a tight mesh that accounts for the structural stability of the extracellular matrix. One of the better-characterised GAG/proteoglycan structures is that of the proteoglycan aggregates found in mammalian cartilage. An example of this is shown in the electron micrograph in figure 1.2, along with a schematic representation of the components of the aggregate.

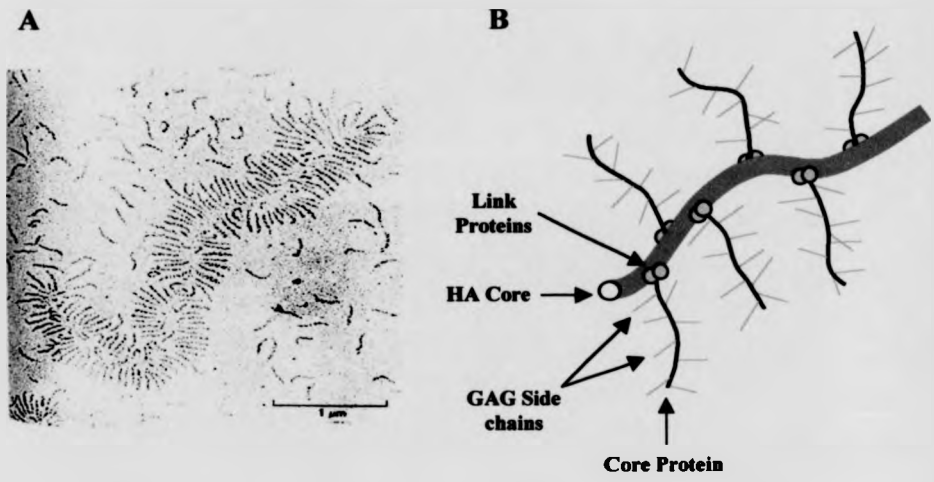


Figure 1.2 – GAG/proteoglycan aggregates found in mammalian cartilage. A: Electron micrograph with platinum shadowing. B: Schematic diagram showing key features of the macromolecule. Both the micrograph and the diagram are taken or adapted from Alberts *et al.* Edition 2, p.807 (Alberts *et al.*, 1989).

As can be seen, the aggregate is composed of proteoglycans bound to a core HA molecule by link proteins. HA differs from the other GAG's in several respects. There are no sulphated residues, the chain length can be very long at around 2-3,000 residues, often reaching molecular weights of several million Daltons. It is not found associated with a core protein and it is not found in branched chains.

The relatively simple structure of HA (as shown in figure 1.3) has lead to the suggestion that this molecule represents an early evolutionary form of GAG (Alberts *et al.*, 1989), whereas the more structurally elaborate species of GAG have evolved as a result of the selective pressure provided by the microbial pathogens which degrade them (Li *et al.*, 2000).

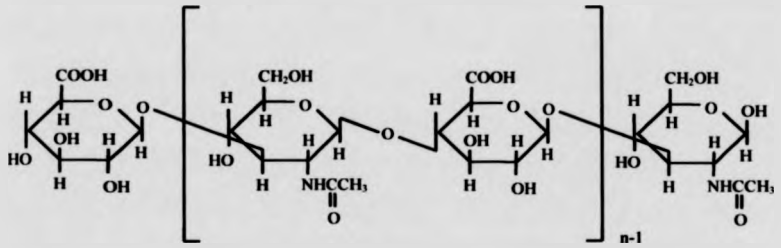


Figure 1.3 – Repeat unit of hyaluronic acid (HA). All substituents are shown.

As can be seen from figure 1.3, HA consists of alternating N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcUA) residues linked by alternating β-1,3 and β-1,4 glucosidic bonds. The repeat unit is termed N-acetylhyalobiuronic acid.

Aqueous solutions containing HA are highly viscous, and increased concentrations of HA can lead to the formation of HA gels. HA-rich fluids act as lubricants and facilitate shock absorption in vertebrate joints. These properties are largely steric in nature; mutual repulsion of the carboxyl groups on the glucuronic acid moieties results in the co-ordination of large numbers of water molecules.

Understanding the roles of HA in the host has advanced over the years from the early appreciation of its importance in tissue structure to implications for tissue differentiation, angiogenesis, metastasis, embryogenesis and tissue repair.

HA is recognised as a major component of the extracellular matrix (West *et al.*, 1985, Reed *et al.*, 1990b, Stamnekovic and Arrufo, 1994, Frost *et al.*, 1996). The distribution of HA varies between tissues and around 50% of total HA is thought to reside in the skin. The turnover of this material is rapid, with 50-75% of tritiated HA injected into rabbit skin metabolised over 24 hours (Reed *et al.*, 1990a)

Several studies have implicated HA in the organisation of tissues undergoing rapid migration and proliferation. These effects are mediated by physical and biochemical factors: The influx of water into HA-rich regions (due to the hygroscopic nature of the molecule) results in host cells being forced apart, allowing for the migration of other cell

types around the region (Toole, 1981). The effects of HA on the cellular environment are also regulated by HA receptors on host cells, including CD44 and RHAMM (Receptor for Hyaluronic Acid Mediated Motility) (Stamnekovic and Arrufo, 1994).

High levels of HA appear to inhibit cellular differentiation in some systems (Kujawa and Tepperman, 1983, Kujawa *et al.*, 1986) including that of the developing chick embryo cornea (Toole, 1981).

HA also plays a complex role in angiogenesis. High molecular mass HA is seen to inhibit formation of vascular tissue, while low mass HA stimulates the process (Feinberg and Beebe, 1983, West *et al.*, 1985). Thus interplay between hyaluronate synthases and hyaluronidases potentiates angiogenesis. This has further implications for various processes ranging from embryogenesis to metastasis.

While the diverse roles played by HA in the host are of evident interest, it is the contribution of HA to ECM structural integrity which is of primary concern when addressing the case of pneumococcal hyaluronate lyase. However, the range of enzymes capable of degrading HA is broad and diverse, as will be discussed below.

1.7.2 – Hyaluronidases degrade HA

As mentioned above, HA is broken down by hyaluronidases and hyaluronate lyases. Together these enzymes form a large and diverse group of proteins from the eukaryotic and prokaryotic kingdoms. The common biochemical function shared by these enzymes is their ability to degrade hyaluronic acid, and as such they are believed to act as spreading factors, albeit in different biological systems. Their genetic and structural diversity is reflected in the way in which these enzymes perform this degradation.

1.7.3 – Mammalian and insect hyaluronidases – PH-20

The mammalian hyaluronidase is an endo-β-N-acetyl-D-hexosaminidase that cleaves between GlcNAc and GlcUA residues to yield primarily tetrasaccharides. The ability of the enzyme to perform transglycosylation reactions then leads to the secondary

production of di- and octa-saccharides. This enzyme is marked by its relatively broad substrate specificity, also being able to degrade chondroitin sulphates (Frost *et al.*, 1996).

This class of hyaluronidase is termed the PH-20 type on the reactivity of one of the earliest examples of this class (from Guinea Pig sperm acrosome) with a particular monoclonal antibody raised against a protein previously termed PH-20 (Primakoff and Myles, 1983). The enzyme from mammalian sperm contributes to the fertilisation process by enabling the sperm to traverse the layer of HA-rich cumulus cells surrounding the egg.

Other PH-20 type hyaluronidases have subsequently been described in insect venom. The enzyme from honey-bee venom was identified on the basis of homology to the PH-20 prototype, and is related to venom hyaluronidases from other hymenopterans. This has lead to the interesting suggestion that the venom spreading factor in these organisms has evolved from the mammalian reproductive system (Frost *et al.*, 1996).

Antibody binding studies suggests that the PH-20 structural class of hyaluronidase may be more widely distributed than has been previously thought. Antibodies raised against the hyaluronidase of *Apis melifera* were shown to be cross reactive with surface proteins shown to be able to bind HA from the spirochaete *Treponema denticola* (Scott *et al.*, 1996).

Homology studies have indicated that the active site of the PH-20 type hyaluronidases lies in the amino terminal portion of the protein. This active site is thought to consist of two conserved acidic residues (glutamic or aspartic acid) and the substrate binding pocket is thought to be lined with hydrophobic residues such as tryptophan and tyrosine (Kreil, 1995).

1.7.4 – Other venom hyaluronidases

Hyaluronidase activities in other venom samples were studied by Cevallos *et al.* (Cevallos *et al.*, 1992) using an in-gel technique which involves PAGE through gels containing HA, renaturation, incubation and visualisation of zones of HA degradation using a modification of the dye-binding technique of Benchetrit *et al.* (Benchetrit *et al.*, 1977) – see section 1.8.2.1, below. Using this technique they were able to identify hyaluronidase activity from the venom of *Dolichovespula maculata* (White faced hornet),

Vespula germanica (Yellow jacket wasp), *Pogonomyrmex rugosus* (Harvester ant), *Heloderma horridum horridum* (Mexican bearded lizard), *Heloderma suspectum suspectum* (Gila monster), *Lachesis muta* (Bushmaster snake), *Crotalus basiliscus* (Mexican west-coast rattle snake), *Bothrops asper* (Central American pit viper), *Micrurus nigrocinctus* (Central American coral snake) and *Centruroides limpidus limpidus* (a scorpion of Mexican origin. In many cases the low intrinsic toxicity of the purified hyaluronidase supports the view that the enzymes contribute to toxicity by facilitating the spread of other venom components (Frost *et al.*, 1996)

1.7.5 – Leech hyaluronidase

The enzyme from the leech (*Hirudo medicinalis*) is an endo-β-D-glucuronidase, which cleaves between GlcUA and GlcNAc residues to yield primarily tetrasaccharides (Yuki and Fishman 1963). It differs from the mammalian enzyme in its absolute substrate specificity, and is of passing interest as a glucuronidase rather than a glucosaminidase, however it has not been extensively characterised and will not be considered further here.

1.7.6 – Bacterial Hyaluronate lyases

The bacterial enzymes comprise another class of HA degrading enzyme. The most extensively characterised are endo-N-acetyl-D-hexosaminidases. Whereas the PH-20 class of enzymes are hydrolases, the bacterial enzymes are lyases, progressively degrading the HA chain by a β-elimination reaction leading to the production of unsaturated disaccharides containing an –ene moiety in the uronic acid component of the molecule. This reaction is illustrated in figure 1.4, below.

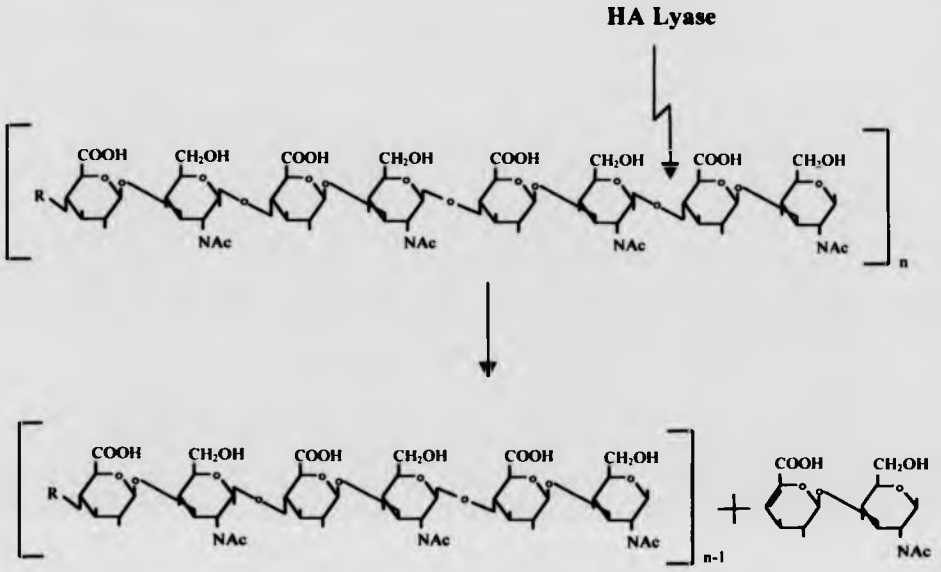


Figure 1.4 – The reaction catalysed by bacterial hyaluronate lyase. Only the key substituents are shown, with the hydroxyl groups simplified to short vertical bars. Note the introduction of a π bond between C4 and C5 of the GlcUA moiety.

The mode of action illustrated in figure 1.4 was first described for the pneumococcal hyaluronate lyase (Linker and Meyer, 1954), and a large number of bacterial polysaccharide lyases have been described, although few in detail. For a review see Sutherland (Sutherland, 1995).

The prokaryotic HA lyases are often cited for their high substrate specificity, although this is not universally true. The HA lyase of the group B Streptococcus, *S. agalactiae* has also been shown to be active against chondroitin sulphate, cleaving at the unsulphated regions of the molecule (Baker *et al.*, 1997, Lin *et al.*, 1997). This wider substrate specificity may reduce trophic restrictions on this organism, of facilitate more effective invasion of host tissues.

1.7.7 – The special case of the hyaluronidases of *Streptococcus pyogenes*

S. pyogenes, the group A streptococcus (or GAS) is one of the limited number of streptococci which produces a capsule containing HA. In *S. pyogenes* HA capsule production has been shown to be directed by the *has* operon (Albertini *et al.*, 1998). Other members of the HA capsule producing group include streptococcal species from Lancefield group C (including isolates *S. dysgalactiae, S. equi* and *S. equi* subspecies *zooepidemicus*) (Gunther *et al.*, 1996).

Interestingly, the authors of this study found that the incidence of production of the HA capsule was very low in clinical human and animal samples. One possible suggestion for this observation is that the HA capsule occludes cell surface adhesins required of adhesion to the host epithelia, however the observation runs counter to the findings of murine infection studies which have associated HA production with virulence (Leonard *et al.*, 1988, Moses *et al.*, 1997). This apparent discrepancy is not easy to resolve, as the sites of isolation of these organisms are not given in the former citation. It is possible that the same play-off between adhesion versus resistance to phagocytosis, modulated by the capsule in pneumococci may operate in the GBS system.

Another suggestion is that the HA capsule may actually contribute to adhesion by attachment to host cells through HA binding to CD44 (Schrager *et al.*, 1998). Furthermore, the notion that co-ordinated expression of the HA capsule contributes to virulence was supported by the identification of a two-component system capable of modulating the expression of the *has* operon (Levin and Wessels, 1998), however the exogenous signal for this system has not yet been identified.

S. pyogenes has been shown to possess a chromosomally located hyaluronidase gene, and is host to at least 2 types of bacteriophage that encode their own hyaluronidase. Given the nature of the GAS HA capsule there is a patent requirement for phage encoded hyaluronidase during the viral infection cycle. The first nucleotide and amino acid sequence of such an enzyme was described by Hynes and Ferretti (Hynes and Ferretti, 1989). This was cloned from the bacteriophage H4489A, and termed HylP. The authors also noted the prior description of an apparently different

hyaluronidase by Benchetrit (Benchetrit *et al,*. 1978) from bacteriophage GT8760. While the restriction fragment profiles of the two viral genomes were identical, the enzymes differed on the basis of their electrophoretic mobility. The authors were unable to conclusively explain the apparent differences between these two hyaluronidases.

The Hynes/Ferretti collaboration then went on to describe another phage-associated hyaluronidase, this time from a T-type 22 strain (Hynes *et al,*. 1995). This was termed HylP2, and was nearly identical to HylP other than the loss of a series of collagen-like G-X-Y repeats. Subsequent analysis of a population of *hylP* type genes demonstrated that these phage encoded enzymes evolve by homologous recombination between strains, not unlike the mechanism of variation driving the evolution of various streptococcal genes (see section 1.3.2 – above) (Marciel *et al,*. 1997).

S. pyogenes also encodes its 'own' hyaluronate lyase, which is distinct from the phage-encoded proteins. This shows homology to other bacterial hyaluronate lyases, notably that of the pneumococcus (53% identity over a 500 residue region). It is notably different from the phage enzyme as it harbours the string of conserved residues which is thought to comprise the active site of the bacterial hyaluronate lyases (see section 1.7.9, below) (Lin *et al,*. 1997).

In their description of the gene encoding HylP2, Hynes *et al.* noted the activity of a 'third hyaluronidase' detected in the extracellular milieu (Hynes *et al,*. 1995). It is suggested here that this secreted enzyme is in fact that encoded by the 'genomic' rather than phage associated locus.

1.7.8 – Phylogenetic associations between known HA degrading enzymes

Some examples of the hyaluronate degrading enzyme classes described above are grouped according to amino acid sequence similarity in the dendrogram shown in figure 1.5. The illustrative purpose of such a dendrogram is twofold: It serves to highlight the genuine relationships between proteins within the same class, and demonstrates the distinction between the classes of HA degrading enzymes.

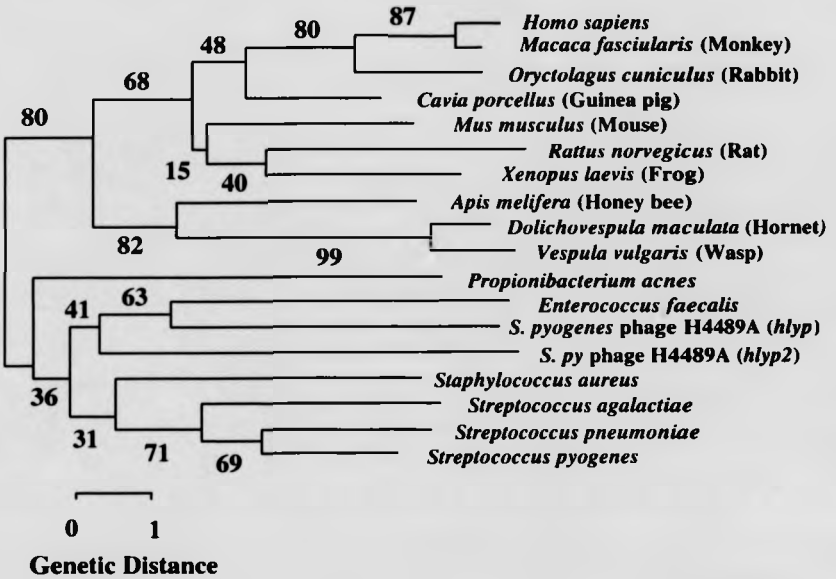


Figure 1.5 - Phylogeny of examples of the major classes of HA degrading enzymes. Alignments of amino acid sequences were made using the Neighbour-Joining method. Bootstraps were calculated on the basis of 500 replicates. The dendrogram was generated using the MEGA suite of programs (Molecular Evolutionary Genetics Analysis, version 1.01, The Pennsylvania State University, University Park, PA 16802).

Several features described in the preceding sections are reflected in the phylogenic tree shown in figure 1.5. The individual clustering of the groups of mammalian and insect hyaluronidases and the prokaryotic hyaluronate lyases is evident, as is the derivation of the insect venom hyaluronidases from the eukaryotic group of enzymes. Some of the bootstrap values are notably low, indicating that the assignment of a given breakpoint at a particular node are less certain than others.

1.7.9 – Conserved features of prokaryotic hyaluronate lyases

The sequences of the bacterial and bacteriophage encoded proteins reveals that a number of residues remain well conserved. These are indicated in the alignment shown in

figure 1.6, as are the absolutely conserved residues thought be the catalytic residues (described in more detail in section 1.7.11).

S. py N – $(X)_{40}$ - **G F Y Q D G S L I D H V V T N A Q S** P L Y K K G I A Y T G A Y G **N V L I D G L**
S. pn N – $(X)_{40}$ - **G F Y Q D G S Y I D H T** ---------- **N V A Y T G A Y G N V L I D G L**
S. ag N – $(X)_{40}$ - **G F Y A D G S Y I D H T** ---------- **N V A Y T G A Y G N V L I D G L**
S. au N – $(X)_{45}$ - **G F Y K D G S Y I D H** ---------- Q ---------- **D V P Y T G A Y G V V L L E G I**
P. ac N – $(X)_{40}$ - **G F R A D G G F I Q H** ---------- S ---------- **H I P Y T G G Y G D V L F S G L**
E.faec N – $(X)_{40}$ - **G L Y P D G S L I Q H G** ---------- Y -------- **F P Y N G S Y G N E L L K G F**

Figure 1.6 – Alignment of the cluster of residues believed to from the active site of the bacterial hyaluronate lyases. Conserved residues are shown bold typeface, catalytic residues are shown in red typeface. Abbreviations, accession numbers (or other reference) and residue positions are as follows; *S. py – Streptococcus pyogenes* (23-101) residues 6003-7715, contig. 265, WIT database sequence (http://www.mcs.anl.gov/wit2), *S. pn – S. pneumoniae* (349-416) L20670, *S. ag – S. agalactiae* (517-584) E1308674, *S. au – Staphylococcus aureus* (241-314) Q59801, *P. ac – Propionibacterium acnes* (222-289) Q59634, *E. faec – Enterococcus faecalis* (437-504) residues 6882-3895, contig. 6165, WIT database sequence.

As can be seen from figure 1.6, 12/28 residues are absolutely conserved within all of the sequences shown. 19/28 (68%) residues are conserved in most of the sequences, with the divergence in the active site region highest for the proteins from *Propionibacterium acnes* and *Enterococcus faecalis*. This reflects the trend shown in the dendrogram in figure 1.5. A clustal alignment of these sequences reveals percentage identity ranging from 47.8% (*S. pneumoniae/S. pyogenes*) to 14.6% (*S. agalactiae/E. faecalis*). As can also be seen, some residues are absolutely conserved among these proteins, three such residues, N349, H399 and Y408 (pneumococcal sequence annotation) will be shown to be critical for enzyme function in section 1.7.11, below. This section describes how structural studies of pneumococcal hyaluronate lyase have helped elucidate the roles of certain residues in the catalytic mechanism of these enzymes.

1.7.10 – First characterisation of hyaluronate lyase from the pneumococcus

The locus encoding the pneumococcal hyaluronate lyase was initially described by Berry *et al.*. (Berry *et al,*. 1994). Using a crude hyaluronate lyase preparation they generated polyclonal mouse antisera, with which they isolated the region harbouring the *hyl* locus from a cosmid library expressing in *E. coli*. Using a combination of restriction digestion and hyaluronate lyase assays (a description of commonly used assays for hyaluronate degradation is given in section 1.8.2.1, below), the region purported to contain the *hyl* locus was subcloned and its sequence derived. This was deposited in the Genbank database under the accession number L20670.

Their open reading frame of 2849 bp encodes a 949 amino acid protein. A strong candidate for a rho-independent transcription terminator was defined about 100 bp 3' to the suggested stop codon. However, poor -35, –10 and ribosome binding elements suggested that the transcriptional start point of the gene was ill defined. This issue will be returned to in section 1.8.6, below.

Expression of the subcloned region in *E. coli*, and subsequent Western analysis indicated that the protein was subject to proteolytic cleavage in the *E. coli* host system, as well as translation from an alternative start codon 489 bp 3' to the predicted start codon. This was not seen in *S. pneumoniae*.

Leaving the nature of the N-terminus of the protein due to the uncertain nature of the assigned promoter elements, the C-terminus contains an excellent candidate for a Gram positive cell surface anchor motif. This motif consists of a hexapeptide with the consensus LPXTG(E) (in this case, LPQTGE) which becomes covalently attached to the stem peptide moiety of cell wall peptidoglycan. This is followed by a stretch of hydrophobic amino acids, which traverse the cell membrane. The terminal structure is a cluster of basic residues which remains associated with the cytoplasm (Navarre and Schneewind, 1999). This strongly suggests that the pneumococcal hyaluronate lyase is associated with the cell-wall.

1.7.11 – Pneumococcal hyaluronate lyase – crystal structure and catalytic mechanism

The earliest indications of which residues might be employed at the catalytic centre of bacterial hyaluronate lyases came from the work of Lin, Averett and Pritchard. Using site directed mutagenesis techniques in *S. agalactiae,* they demonstrated that a key histidine residue is vital for enzymatic activity (Lin *et al,*. 1997). This residue corresponds to the histidine shown in red in figure 1.6.

During the past 2-3 years the structural biology group of Mark Jedrzejas have been interested in elucidating the crystal structure and catalytic mechanism of pneumococcal hyaluronate lyase. These efforts have culminated in detailed crystal structures, resolved to 1.56Å in the absence of substrate, and 1.7 Å in the presence HA disaccharide repeat units (Jedrzejas *et al,*. 1988, Li *et al,*. 2000, Ponnuraj and Jedrzejas, 2000). Their proposed catalytic mechanism is supported circumstantially by the conservation of key residues within this class of enzyme (as described above) and by site-directed mutagenesis studies. Their work is summarised as follows:

Their earliest attempts at crystallising the protein resulted in stability problems, and lead them to adopt a strategy where the expressed protein was truncated, lacking the N-terminal 163 residues (from the sequence deposited under the accession number L20670) (Berry *et al,*. 1994, Jedrzejas *et al,*. 1988). Subsequent work with this truncated construct resulted in the derivation of stable crystals, and high resolution analysis of the structure. The protein was shown to consist of 2 structural domains, an N-terminal α-helical region and a C-terminal β-sheet rich domain, with a prominent cleft between them. The overall dimensions of the protein are 59 X 59 X 88 Å. These features are shown in figure 1.7.

Figure 1.7 – Domain structure of pneumococcal hyaluronate lyase. A: Ribbon diagram of the α-domain, B: Ribbon diagram of the β-domain, C: Surface electrostatic map of the complete crystal structure. Positive charges are indicated in blue, negative in red. Both images are taken from the paper of Li *et al.* (Li *et al,*. 2000).

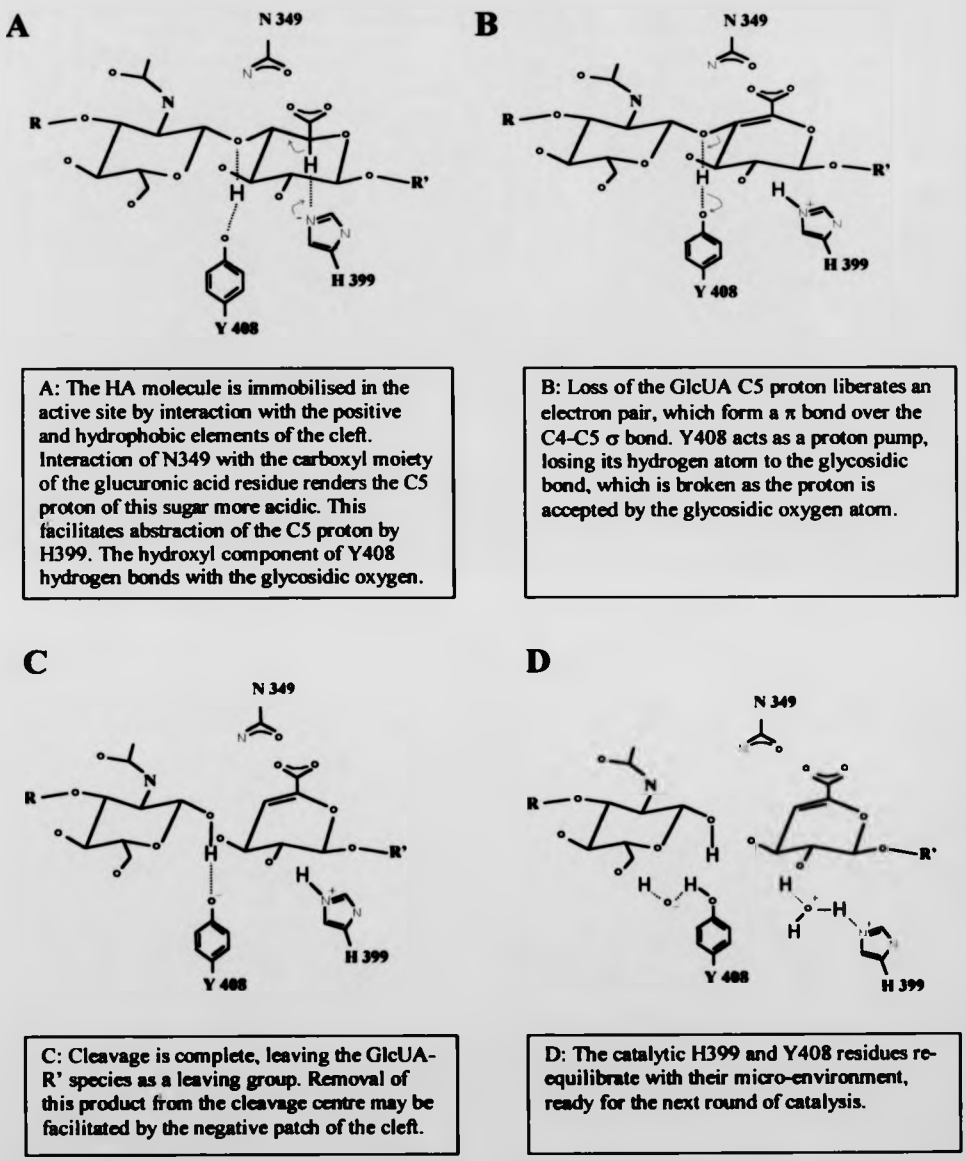
The structure revealed in this analysis shows several key features; The β-domain consists of a 4-layered sandwich, and is not thought to contribute significantly to catalysis. The function of this domain is suggested to be the modulation of catalytic activity upon binding small molecules or cations, such as Ca^{2+} (Li *et al,*. 2000).

The α-domain contains the catalytic residues and has the conformation of a twisted α/α barrel. There are also some structures which partially occlude the cleft between the domains in the region of the linker peptide. This feature will be returned to below.

The cleft itself binds the HA molecule and contains 4 key features; a positively charged region, responsible for primary interaction with its polyanionic substrate. A

patch of aromatic residues interacts with a stretch of C-H features on the substrate molecule (distributed over three successive sugar rings). Modelling of the substrate into the cleft suggests that this patch is required for accurately positioning the substrate such that the glycosidic bond to be cleaved is selected and held in the appropriate position at the active site. The third key region is the active site itself, and is described below. The fourth region is a small negatively charged region, which lies at the 'post-catalytic' end of the cleft. This may be required for electrostatic repulsion of the disaccharide product after it is cleaved from the rest of the HA chain.

The mechanism of catalysis is thought to involve a number of residues, and interactions between the enzyme and substrate, however this description will be limited to the roles of three key residues, N349, H399 and Y408. Their term for the type of reaction shown is 'Proton Acceptance and Donation' or PAD.



A: The HA molecule is immobilised in the active site by interaction with the positive and hydrophobic elements of the cleft. Interaction of N349 with the carboxyl moiety of the glucuronic acid residue renders the C5 proton of this sugar more acidic. This facilitates abstraction of the C5 proton by H399. The hydroxyl component of Y408 hydrogen bonds with the glycosidic oxygen.

B: Loss of the GlcUA C5 proton liberates an electron pair, which form a π bond over the C4-C5 σ bond. Y408 acts as a proton pump, losing its hydrogen atom to the glycosidic bond, which is broken as the proton is accepted by the glycosidic oxygen atom.

C: Cleavage is complete, leaving the GlcUA-R' species as a leaving group. Removal of this product from the cleavage centre may be facilitated by the negative patch of the cleft.

D: The catalytic H399 and Y408 residues re-equilibrate with their micro-environment, ready for the next round of catalysis.

Figure 1.8 – Reaction catalysed at the catalytic centre of pneumococcal hyaluronate lyase. This figure is adapted from *et al.* (Li *et al,*. 2000), with some elaborations. Key electron pair transitions are indicated by double headed red arrows.

Figure 1.8 describes the catalytic mechanism proposed by Li *et al.* The stabilisation of the substrate by N349 is brought about by hydrogen bonding via environmental water molecules. This has an electron withdrawing effect on the GlcUA C5 atom. Not unlike the ruse of a pickpocket, while the 'attention' of the C5 atom is diverted by N349, H399 is able to remove the proton from the other plane of the molecule. This unusual mechanism neatly explains the sequence of events which take place thereafter, leaving the novel unsaturated disaccharide product.

The absolute conservation of N349, H399 and Y408 across this group of enzymes supports their hypothesis, however, site directed mutagenesis approaches were taken to confirm and evaluate catalytic attenuation. N349A, H399A and Y408F mutants showed 6, 12 and 0% of wild type activity. The residual activity of the H399A mutation was attributed to the remaining function of N349 in making the GlcUA C5 carbon more acidic, allowing its proton to be lost without having a specific residue being reduced.

One or two implications for the evolution of these enzymes have also arisen from the structural work described above; The ability of bacteria to degrade GAG's of higher organisms clearly affords them a toehold towards pathogenesis, and both groups of organisms degrade HA via quite different means. The striking difference in the mechanism of HA degradation between these classes of organisms may reflect the evolutionary pressure of such competition, and the higher specific activity of the bacterial enzymes may indicate their 'response' to the barrier functions of mammalian HA turnover (Li *et al,*. 2000).

Another interesting point concerns the relationship of the pneumococcal hyaluronate lyase fold structure to other degradative enzymes. The barrel structure of the α-domain resembles structures seen in glycoamylase and alginate lyase, and the complete structure (as determined by the authors) bears striking resemblance to that of chondroitin AC lyase from *Flavobacterium heparinum* (although the latter enzyme has a wider opening to the cleft, which almost certainly accounts for its wider substrate specificity (Fethiere *et al.*, 1999)). The implication being that the structure and function of the pneumococcal enzyme may provide a paradigm for those of other proteins involved in the degradation of simple and complex polysaccharide substrates. It will be interesting to see whether the PAD mechanism of catalysis manifests itself in other degradative systems.

1.8 – The work presented in this thesis – brief rationale and methodology

An outline of the aims and approaches taken in this body of work are briefly described below. A broader explanation of the basis of each area of investigation is given at the beginning of each major section within chapter 3.

1.8.1 – Variation at the locus encoding hyaluronate lyase, *hyl*

As described in section 1.3.2, numerous pneumococcal genes are known to evolve by recombination between homologous stretches of DNA. This will be of interest in addressing the nature of the evolution of pneumococcal hyaluronate lyase. Also, consideration of Hyl as a candidate for inclusion in next generation polysaccharide conjugate vaccines must necessarily include an appraisal of the genetic, and potentially antigenic variation at the *hyl* locus.

As such it is necessary to analyse the *hyl* locus from a wide range of strains, which differ in geographic and clinical sites of isolation, serotype and year of isolation, to ensure that a sample is taken which is representative of the population as a whole.

The approach taken in our laboratory to conduct sequence variation analyses using a large number of strains is that of High Resolution Restriction Analysis (HRRA), sometimes referred to as Restriction Fragment Length Polymorphism analysis (RFLP).

This technique is described in detail in chapter 2, section 2.2.1. In brief, it enables the identification of distinct allelic variants and the analysis of large numbers of individual loci without the economic burden of sequencing the entire locus. The 'window' of the restriction endonuclease recognition footprint along with a statistical inference algorithm can be used to estimate the overall percentage divergence between loci with different restriction digest profiles. Representative alleles of key restriction profiles are then characterised by direct sequence analysis.

Previous studies have focused on the distribution of genes within members of clonal groups of pneumococci specifically associated with disease (Coffey *et al,*. 1991, Sibold *et al,*. 1992, Enright and Spratt, 1998). A recent study has addressed the occurrence of strains with identical genetic backgrounds in invasive disease and carriage,

and shown that members of clones frequently associated with invasive disease are also found in the carried state (Muller-Graf *et al,*. 1999). A similar analysis was undertaken in the work described here to investigate the potential association of particular *hyl* allelic variants with carriage, incidence of pneumonia, meningitis, septicaemia and otitis media. This section also describes screening for homologous genes in other streptococci.

1.8.2 – Phenotypic analysis of hyaluronate lyase production

In 1995 Kostyukova *et al.* published work which implied that hyaluronate lyase activity from strains isolated from the cerebrospinal fluid (CSF) of meningitis patients was significantly higher than that of strains isolated from asymptomatic carriers of *S. pneumoniae*, indicating a relationship between hyaluronate lyase phenotype and site of isolation (Kostyukova *et al,*. 1995). However, it was not clear from this work whether any rigorous attempts had been made to exclude large numbers of clonally derived isolates from their sample. Indeed, the invasive isolates were collected from a paediatric clinical environment, just the sort of setting one might be expected to find predominant pneumococcal clones (see section 1.4). It was therefore decided to re-investigate the relative activity of hyaluronate lyase using strains from a wide range of disease types and carriage, with an emphasis placed on avoiding multiple samples of clonal isolates during collection of the strains to be assayed.

The analysis of phenotype requires the use of the appropriate assay(s). The section below describes the methods that are commonly available.

1.8.2.1 –Hyaluronate lyase assays

These measure either a decrease in the concentration of substrate, or an increase in the concentration digest products. The key assays available are described below.
One of the earliest assays used to determine the concentration of hyaluronic acid is based on the highly viscous nature of hyaluronic acid solutions (Hass, 1946). Changes in the viscosity of such solutions is indicative of the conversion of high molecular weight hyaluronic acid polymers to the shorter oligosaccahrides, tetrasaccharides and

disaccharides which result from the action of the various hyaluronidases. This assay was successfully used for a time, but has been described as a process which is 'tedious, time consuming, and requires large amounts of hyaluronic acid' (Dorfman and Ott, 1947).

Another early assay is the mucin clot formation prevention assay. It was shown by (McClean and Hale, 1941) that, in the presence of albumin, hyaluronic acid precipitates out of solution upon acidification forming a visible clot. The enzymatic degradation of hyaluronic acid prior to performing this reaction prevents the formation of this clot, and was used as a method for the estimation of hyaluronate lyase activity.

This method was modified in the turbidimetric method of Seastone and Kass (Seastone, 1939, Kass and Seastone, 1944), who found the viscosity measurement assay 'inconvenient'. They demonstrated that clot formation was preceded by the formation of a turbid suspension upon addition of a limiting amount of hyaluronate lyase to albumin/hyaluronate solutions prior to acidification. However this method has a number of drawbacks; it is not highly quantitative between experiments, it is effective over a limited concentration range, and the observation that some hyaluronate preparations require pre-digestion with bovine testicular hyaluronidase in order to form the desired characteristic turbid suspension.

The hyaluronic acid determination assays described above were originally used in the investigation of mammalian hyaluronic acid and hyaluronidases. Precipitation of acidified hyaluronate was also used as the basis of the plate assay of Smith and Willet (Smith and Willet, 1968), who were specifically interested in the production of bacterial glycosidases. In their assay, hyaluronic acid and bovine serum albumin (BSA) were incorporated into agar plates, on which various cultures of micro-organisms were grown. After growth and expression of the bacterial hyaluronate lyases, the plates were flooded with acetic acid. The high molecular hyaluronic acid/albumin complexes were precipitated, leading to an opaque or cloudy appearance. Around some colonies, clear zones were seen. This was taken to indicate that the hyaluronic acid was degraded by soluble bacterial hyaluronate lyases. The relative sizes of the zones of clearance could be taken as a measure of the relative amount of enzyme produced, although it could also reflect the different solubility characteristics of different hyaluronate lyases.

During the early stages of this work, this assay was considered. However, after discussion with collaborators working in a related field (K. Homer and D. Beighton, *personal communication*), it was decided that this assay would not be appropriate. The work done by Homer and Beighton with *S. intermedius* showed the assay to have a number of drawbacks, of which two were seen to be significant: Strains which were hyaluronate lyase positive in the 'Stains-all' assay (described below) were only weakly positive in the plate assay. Also, as the technique requires co-precipitation of hyaluronic acid with BSA, the degradation of the latter by proteases would also lead to a positive result. Hence hyaluronate lyase activity could be masked by the activity of other bacterial enzymes. This lack of assay specificity was considered to be undesirable.

Levy and McAllan demonstrated that it was possible to measure the concentration of acetylated hexosamines in a colourimetric reaction with 4-(N,N-dimethylamino)-benzaldehyde (Levy and McAllan, 1959). This assay has been used to measure the activity of streptococcal glycosidases by measuring the accumulation of digest products (Homer *et al.*, 1994) and was shortlisted as a suitable assay for use in the work described here.

Benchetrit and coworkers described a quite different type of assay (Benchetrit *et al.*, 1977). Theirs was a colourimetric assay based on thiazolium compound 1-ethyl-2-[3-(1-ethylnaptho[1,2*d*]thiazolin-2-ylidine)-2-methylpropenyl]naptho[1,2*d*]thiazolium bromide (marketed by Sigma-Aldrich Chemicals as 'Stains-all').

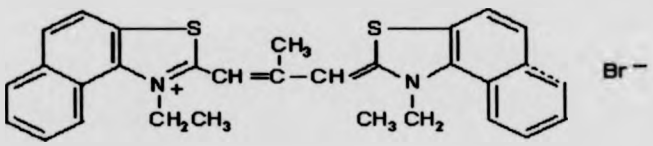


Figure 1.9 – Chemical structure of 1-ethyl-2-[3-(1-ethylnaptho[1,2*d*]thiazolin-2-ylidine)-2-methylpropenyl]naptho[1,2*d*]thiazolium bromide ('Stains-all').

This is a carbocyanine dye, which interacts specifically with hyaluronic acid to give rise to a characteristic absorbance peak in the visible spectrum (650nm). This is seen as a change in colour of the solution from deep purple to blue, and can be measured spectrophotometricaly. This interaction is specific for high molecular weight hyaluronic acid, and not the products of digestion by hyaluronate lyase, making it possible to monitor the degradation of substrate in hyaluronate lyase assays.

Another assay for hyaluronate lyase activity was described by Pritchard and co-workers (Pritchard *et al.*, 1994). This assay is based on the formation of an unsaturated disaccharide as a major product of hyaluronate digestion by the bacterial hyaluronate lyases (see section 1.7.9). The π bond at the 4-5 position of the glucuronic acid moiety of the disaccharide leads to detectable absorbance in the U.V/vis spectrum. The assay described here is based on increased absorbance at 232nm as the hyaluronate lyase reaction proceeds. This is depicted in figure 1.10.

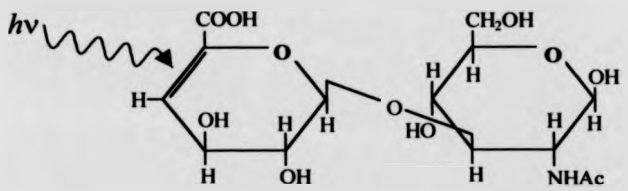


Figure 1.10– Structure of the disaccharide produced by pneumococcal hyaluronidase, a derivative of N-acetylhyalobiuronic acid. Note the presence of the pi bond in the glucuronic acid moiety.

This assay has the advantage that it is straightforward to perform, on readily available equipment, and it can be performed without the introduction of costly reagents.

1.8.3 – Hyaluronate lyase knockout mutagenesis

The aim of this area of work is to establish that hyaluronate degradation by the pneumococcus is solely attributable to the action of hyaluronate lyase, and that there is no functional redundancy with regard to this activity.

The analysis of a naturally occurring deletion mutant is undertaken, as well as the construction of a defined mutant by interposon mutagenesis.

This feeds into the next section, which is specifically concerned with the ability of pneumococci to grow on HA as a sole carbon source.

1.8.4 – HA utilisation by *S. pneumoniae*

Hyaluronidases and hyaluronate lyases are traditionally considered to act as spreading factors, and thus contribute to pathogenesis. Since HA is a potentially rich carbon source and is in abundance in those sites at which the pneumococcus is found, it is tempting to

speculate that hyaluronate lyase might contribute to pathogenesis by providing the pneumococcus with a valuable source of metabolisable carbon.

This is analysed using chemically defined media, to which various carbon sources have been added. The results of these enquiries are given in chapter 3, section 3.4.

1.8.5 – Effect of capsule on hyaluronate lyase phenotype

As mentioned in section 1.5.1, variation in the amount of capsule produced by the pneumococcus potentiates its response to the host in different environments (being inhibitory to both adhesion and phagocytosis). The manner in which hyaluronate lyase production fits into this process of variation. Selection of acapsular mutants was achieved using a biofilm growth model.

1.8.6 – Hyaluronate lyase – upstream sequence elements

As indicated in section 1.7.10, the hyaluronate lyase sequence as described by Berry *et al.* is very likely to be incomplete at its 5' extremity (Berry *et al.*, 1994). At least three lines of evidence support this. The first is the unsatisfactory nature of the ascribed promoter elements, the –35 and –10 sequences are divergent from the consensus sequences of these elements, and the Shine-Dalgarno sequence consists of a single pair of guanine residues. On the other hand, there is a C-terminal anchor motif in the *hyl* translation product which matches the consensus for this motif perfectly. The strong indication that this is a protein which is cell surface expressed, for at least some of the time, then requires that the protein possess some means of trafficking to this location. This would normally be in the form of a Gram positive N-terminal signal sequence. Such a feature is not present in the sequence as given.

The third indication that the Berry sequence is incomplete comes from the discovery by Gase *et al.* that the related enzyme, from *S. agalactiae* is indeed longer than was previously thought, and shows the features of a classical N-terminal cleaveable signal sequence (Gase *et al.*, 1998).

2.1 – Materials

2.1.1 - Growth Media

Liquid cultures of *S. pneumoniae* were grown in either BHI (brain heart infusion) broth or pneumococcal C-Medium. BHI (Difco) was made up in dH_2O at 37g l^{-1}.

C-medium is derived from the original formulation of Lacks *et al.* (Lacks and Hotchkiss 1960). Cultures were grown at 37°C without shaking.

C-Medium

Pre-C; 160ml
Supplement; 5.2ml
Glutamine [5mgml^{-1}]; 0.8ml
AdamsIII; 4ml
Pyruvate (2%w/v); 2ml
Phosphate buffer [1M]; 6ml
Yeast Extract (5%w/v); 3.6ml
Filter sterilised prior to storage at 4°C.

Composition of the individual components;

Pre-C

Sodium acetate (anhydrous); 1.23 g l^{-1}
Casamino Acids (Difco – Vitamin free, acid hydrolysed); 5 g l^{-1}
L-Tryptophan; 5 g l^{-1}
L-Cysteine 50 mg l^{-1}
pH titrated to 7.4-7.6 with 10M NaOH

Supplement

3 in 1 Salts; 15ml
Glucose (20% w/v); 30ml
Sucrose (50% w/v); 1.5ml
Adenosine [2mgml^{-1}]; 15ml
Uridine [2 mgml^{-1}]; 15ml

3 in 1 Salts

$MgCl_2.6H_2O$; 100 g l^{-1}
$CaCl_2$ (anhydrous); 500 mg l^{-1}
$CaCl_2.2H_2O$; 660 mg l^{-1}
$MnSO_4.4H_2O$ [0.1M]; 0.2 ml l^{-1}

Phosphate Buffer (gives final pH = 8.0)

KH_2PO_4 [1M]; 26.5ml
K_2HPO_4 [1M]; 473.5ml

Adams III
Adams I; 48ml
Adams II; 12ml
Asparagine; 600mg
Choline chloride; 60mg
$CaCl_2$ [0.1M]; 480µl

Adams II
$FeSO_4.7H_2O$; 500mg l^{-1}
$CuSO_4.5H_2O$; 500mg l^{-1}
$ZnSO_4.7H_2O$; 500mg l^{-1}
$MnCl_2.4H_2O$; 200mg l^{-1}

Adams I

Biotin [2mg ml^{-1}]; 7.5µl
Nicotinic acid; 15mg
Pyridoxine hydrochloride; 17.5mg
Calcium pantothenate; 60mg
Thiamine hydrochloride; 16mg
Riboflavin; 7mg
All solutions were filter sterilised and stored at 4°C.

Modified C-Medium

The experiments described in chapter 3, section 3.4 involve the use of C-medium containing defined carbohydrates. For this, the components were used as described above, with the exception that supplement was made without glucose or sucrose, but contained either no additional carbohydrate, or HA, GlcNAc, GlcUA or glucose added to 0.5mg ml^{-1} (molar equivalents were not used as the experiments were qualitative rather than quantitative).

E.coli was cultured in LB (Luria-Bertani) medium. Liquid cultures were incubated at 37°C with shaking at 220rpm.

LB Medium

Tryptone; 10g l^{-1}
Yeast extract; 5g l^{-1}
NaCl; 10g l^{-1}
pH adjusted to 7 with 5N NaOH

Solid media

S. pneumoniae was grown on BHI/blood agar (14.8g agar (Becton-Dickinson Microbiology Systems) per l BHI) supplemented with 20ml sterile defibrinated sheep blood (TCS Microbiology). *E.coli* was cultured on LB agar (14.8g agar per l LB). Plates were incubated at 37°C under 5%CO_2/95% atmospheric air.

Antibiotic selection

Selection was made on chloramphenicol (crm) at a concentration of 30μg ml^{-1} for *E. coli* and 5μg ml^{-1} for *S. pneumoniae*.

2.1.2 Pneumococcal Strains

The pneumococcal strains used in these studies are listed in table 2.1, with information given where available. All strains are stored in BHI/Glycerol (15%v/v) at –80°C. Temporal distibution, country/site of isolation and serotype were varied to select against the inclusion of multiple clonal isolates. REP PCR was employed to confirm that strains were genotypically distinct, giving rise to diverse band patterns (data not shown).

Strain	Strain ID	Serotype	Country	Year	Site of Isolation
R6	359	ns*	USA	Ca. 1930	CSF
141928	885	1	Kenya	1991	Throat
12B-13E	922	3	(UK)Oxford	1994	Throat
12B-24G	948	4	(UK)Oxford	1994	Throat
142144	889	5	Kenya	1990	Throat
12B-10E	914	6A	(UK)Oxford	1994	Throat
13B-16K	949	6A	(UK)Oxford	1995	Throat
12B-18G	954	6B	(UK)Oxford	1994	Throat
12B-11F	956	6B	(UK)Oxford	1994	Throat
86013	873	8	Kenya	1990	Throat
86033	877	9L	Kenya	1990	Throat
12B-4D	965	9N	(UK)Oxford	1994	Throat
2248	864	9V	Kenya	1990	Throat
141926	884	10A	Kenya	1991	Throat
12B-7C	902	10A	(UK)Oxford	1994	Throat
86034	878	10F	Kenya	1990	Throat
132209	882	10F	Kenya	1990	Throat
13B-27J	903	11A	(UK)Oxford	1995	Throat
85999	869	12F	Kenya	1990	Throat
100497	880	13	Kenya	1990	Throat
55610	862	14	Kenya	1991	Throat
13B-22J	904	14	(UK)Oxford	1995	Throat
86003	872	15A	Kenya	1990	Throat
12B-28G	905	15A	(UK)Oxford	1995	Throat
12B-21E	901	15B	(UK)Oxford	1994	Throat
12B-4F	906	15C	(UK)Oxford	1994	Throat
86029	875	16F	Kenya	1990	Throat
13B-23C	907	16F	(UK)Oxford	1995	Throat
55616	863	17F	Kenya	1991	Throat
12B-34C	908	17F	(UK)Oxford	1995	Throat
86030	876	18A	Kenya	1990	Throat
86002	871	18C	Kenya	1990	Throat
12B-33B	909	18C	(UK)Oxford	1995	Throat

12B-33B	909	18C	(UK)Oxford	1995	Throat
13B-15F	910	19A	(UK)Oxford	1995	Throat
13B-12L	911	19F	(UK)Oxford	1995	Throat
12B-10D	913	19F	(UK)Oxford	1994	Throat
13B-8K	916	20	(UK)Oxford	1995	Throat
12B-11I	917	21	(UK)Oxford	1994	Throat
12B-11H	957	21	(UK)Oxford	1994	Throat
13B-7I	918	22F	(UK)Oxford	1995	Throat
12B-25C	952	22F	(UK)Oxford	1994	Throat
13B-9C	943	23A	(UK)Oxford	1995	Throat
29636	861	23B	Kenya	1991	Throat
12B-23I	919	23F	(UK)Oxford	1994	Throat
12B-14C	921	23F	(UK)Oxford	1994	Throat
12B-8J	923	23F	(UK)Oxford	1994	Throat
85991	866	25	Kenya	1990	Throat
85994	867	29	Kenya	1990	Throat
100493	879	32A	Kenya	1990	Throat
13B-11J	945	33F	(UK)Oxford	1995	Throat
12B-5L	967	35F	(UK)Oxford	1994	Throat
12B-21H	946	37	(UK)Oxford	1994	Throat
12B-28J	947	38	(UK)Oxford	1995	Throat
E/226	355	14	Uruguay	1995	Blood
619	53	1	Spain	1987	Vagina
141950	322	3	Kenya	1991/92	Sputum
102	393	4	-	-	-
964	39	4	Spain	1988/89	Blood
NSM2	187	-	Spain	1989	Sputum
GM133	168	6	Spain	1989	Wound
GM134	169	6	Spain	1989	Wound
4880/89	194	6B	Spain	1989	Eye
132206	312	7	Kenya	1991/92	Sputum
142145	339	7	Kenya	1991/92	Sputum
GM62	123	7	Spain	1988/89	LRT
555	31	8	Spain	1988/89	Blood
GM21	91	8	Spain	1988/89	Blood
85983	310	9	Kenya	1991/92	Pernasal
181	3	9V	Poland	1994/95	Throat
142163	324	10	Kenya	1991/92	Blood
GM14	87	10	Spain	1988/89	Blood
142162	329	10	Kenya	1991/92	Blood
86007	303	11	Kenya	1991/92	Blood
GM39	103	11	Spain	1988/89	LRT
142154	326	12	Kenya	1991/92	Blood
132217	340	12	Kenya	1991/92	Blood
549	30	13	Spain	1988/89	Blood
651	59	13	Spain	1988	Ear
4473/90	200	-	Spain	1990	Blood
7447/90	201	-	Spain	1990	Blood
9058/90	213	-	Spain	1990	Ear
8547/93	231	-	Spain	1993	Peritoneal Fluid
544	28	15	Spain	1988/89	Sputum
1677/94	237	-	Spain	1994	Eye
100494	308	16	Kenya	1991/92	Sputum

631	56	16	Spain	1987	Vagina
100496	319	16	Kenya	1991/92	Sputum
GM123	162	17	Spain	1988/89	LRT
100522	321	18	Kenya	1991/92	Blood
GM1	75	18	Spain	1988/89	Blood
132224	327	18	Kenya	1991/92	Blood
86012	307	19	Kenya	1991/92	Throat
521	26	19	Spain	1988/89	Sputum
85997	315	19	Kenya	1991/92	Sputum
141954	333	20	Kenya	1991/92	Blood
GM122	161	20	Spain	1988/89	Blood
4875/89	256	-	Spain	1989	Peritoneal Fluid
GM66	125	21	Spain	1988/89	Other
100520	318	23	Kenya	1991/92	Blood
GM67	126	23	Spain	1988/89	Blood
E/208	352	23F	Uruguay	1994	Pleural Fluid
18/10/71	379	24	Papua	1971	-
GM120	159	34	Spain	1988/89	Other
7026/70	392	35	-	-	-
GM92	145	3	Spain	1988/89	Blood
110K/70	381	42	Papua	1970	
PN TYPE 45	398	45	UK	-	-
WU2	470	3	UK	1995	-
she15	485	3	UK	1995	-
she10	483	3	UK	1995	Ear
too6	481	3	UK	1995	Ear
gbo5	476	3	UK	1995	-
pn3	472	3	UK	1995	-
SPOI	471	3	UK	1995	-
ply19	492	3	UK	1995	Ear
ply12	490	3	UK	1995	Ear
new1	489	3	UK	1995	-
wol10	487	3	UK	1995	-
sto8	482	3	UK	1995	Ear
por4	480	3	UK	1995	Ear
nor16	479	3	UK	1995	Ear
lei17	478	3	UK	1995	Ear
lei29	477	3	UK	1995	Ear
lei17	475	3	UK	1995	Ear
she23	473	3	UK	1995	-
mid2	466	3	UK	1995	Ear
D39	1001	3	UK	1995	-
ply22	474	3	UK	1995	Ear
0100993	860	3	UK	1995	-
gui5	491	3	UK	1995	Ear
141950	322	3	Kenya	1991	Blood
liv14	494	3	UK	1995	-
409	495	3	UK	1993	Horse
7731	496	3	UK	1992	Horse
600142	497	3	UK	1996	Horse
3307	498	3	UK	1992	Horse
4826	499	3	UK	1992	Horse

5530	501	3	UK	1992	Horse
21.2	502	3	UK	1992	Horse
5	503	3	Ireland	1987	Horse
15	504	3	Ireland	1987	Horse
17	505	3	Ireland	1987	Horse

Table 2.1 – Strains used in genetic analysis.
*ns = non-serotypable (acapsular).

Strain	Strain ID	Serotype	Country	Year	Site of Isolation
792	36	-	Spain	1989	Blood
GM5	77	-	Spain	1989	Blood
GM8	81	15	Spain	1989	Blood
GM43	107	16	Spain	1989	Blood
GM47	110	13	Spain	1989	Blood
GM57	116	14	Spain	1989	Blood
GM58	117	11	Spain	1989	Blood
85997	314	14	Kenya	1992	Blood
E206	342	14	Uruguay	1994	Blood
E213	353	6B	Uruguay	1994	Blood
E228	354	6B	Uruguay	1994	Blood

Table 2.2 - Additional blood isolates used in phenotypic analysis.

Strain	Strain ID	Serotype	Country	Year	Site of Isolation
GM131	166	4	Spain	1989	LRT
GM132	167	13	Spain	1989	LRT
159	432	16	-	1994	Sputum
85990	304	19	Kenya	1992	Sputum
GM108	154	3	Spain	1989	LRT
GM145	176	23	Spain	1989	LRT
100509	301	14	Kenya	1992	Sputum
100506	317	14	Kenya	1992	Sputum
nor1	467	3	UK	1995	Sputum
ash	468	3	UK	1995	Sputum
484	27	6	Spain	1989	Sputum
1177	44	9	Spain	1989	Sputum
1182	45	14	Spain	1989	Sputum
1183	46	6	Spain	1989	Sputum
GM42	106	9	Spain	1989	Sputum
GM84	139	6	Spain	1989	Sputum
GM124	163	14	Spain	1989	Sputum

Table 2.3 - Additional lung isolates used in phenotypic analysis.

Strain	Strain ID	Serotype	Country	Year	Site of Isolation
664	63	6B	Spain	1986	Ear
48/90	197	-	Spain	1990	Ear
1139/89	192	-	Spain	1989	Ear
273	10	14	Poland	1995	Ear
667	65	23	Spain	1988	Ear
630	55	16	Spain	1987	Ear
E159	347	14	Uruguay	1994	Ear
ply19	492	-	UK	1995	Ear
7597	425	-	-	1994	Ear
68961	430	-	-	1994	Ear
5408	428	-	-	1994	Ear

Table 2.4 - Additional otitis isolates used in phenotypic analysis.

Strain	Strain ID	Serotype	Country	Year	Site of Isolation
-	1795	23F	Poland	-	CSF
-	1796	14	Poland	-	CSF
-	1797	38	Poland	-	CSF
-	1798	31	Poland	-	CSF
-	1799	18B	Poland	-	CSF
-	1800	4	Poland	-	CSF
-	1801	1	Poland	-	CSF
-	1802	14	Poland	-	CSF
-	1803	35B	Poland	-	CSF
-	1804	9N	Poland	-	CSF
-	1805	19F	Poland	-	CSF
-	1806	11A	Poland	-	CSF
-	1807	12F	Poland	-	CSF
-	1808	A	Poland	-	CSF
-	1809	16F	Poland	-	CSF
-	1810	2	Poland	1996	CSF
-	1811	1	Poland	1996	CSF
-	1812	2	Poland	1996	CSF
-	1813	18C	Poland	1996	CSF
-	1814	9V	Poland	1996	CSF
-	1815	19A	Poland	1996	CSF
-	1816	3	Poland	1996	CSF
-	1817	7F	Poland	1996	CSF
-	1818	1	Poland	1996	CSF
-	1819	3	Poland	1996	CSF
-	1820	9N	Poland	1996	CSF
-	1821	4	Poland	1996	CSF
-	1822	1	Poland	1996	CSF
-	1823	14	Poland	1996	CSF
-	1824	7F	Poland	1996	CSF
-	1825	12F	Poland	1996	CSF
-	1826	1	Poland	1996	CSF
-	1827	22F	Poland	1996	CSF
-	1828	12F	Poland	1996	CSF
-	1829	9V	Poland	1996	CSF

-	1830	6B	Poland	1997	CSF
-	1831	15C	Poland	1997	CSF
-	1832	1	Poland	1997	CSF
-	1833	8	Poland	1997	CSF
-	1834	23F	Poland	1997	CSF
-	1835	6B	Poland	1997	CSF
-	1836	7F	Poland	1997	CSF
-	1837	33F	Poland	1997	CSF
-	1838	8	Poland	1997	CSF
-	1839	8	Poland	1997	CSF
-	1840	22F	Poland	1997	CSF
-	1841	34	Poland	1997	CSF
-	1842	3	Poland	1997	CSF
-	1843	18C	Poland	1997	CSF
-	1844	23F	Poland	1997	CSF
-	1845	23F	Poland	1997	CSF
-	1846	-	Poland	1997	CSF
-	1847	-	Poland	1997	CSF
-	1848	-	Poland	1997	CSF
-	1849	-	Poland	1997	CSF
-	1850	-	Poland	1997	CSF
-	1851	-	Poland	1997	CSF
-	1852	-	Poland	1997	CSF
-	1853	-	Poland	1998	CSF
-	1854	-	Poland	1998	CSF
-	1856	-	Poland	1998	CSF
-	1857	-	Poland	1998	CSF
-	1858	-	Poland	1998	CSF
-	1859	-	Poland	1998	CSF
-	1860	-	Poland	1998	CSF
-	1861	-	Poland	1998	CSF
-	1862	-	Poland	1998	CSF
-	1863	-	Poland	1998	CSF
-	1864	-	Poland	1998	CSF
-	1865	-	Poland	1998	CSF
-	1866	-	Poland	1998	CSF
-	1867	-	Poland	1998	CSF
-	1868	-	Poland	1998	CSF

Table 2.5 - Additional meningitis isolates used in phenotypic analysis.

Strain	Strain ID	Other Name	Species	Group*	Site of Isolation
Col 1	667	PN92/1207	*S. pneumoniae*		na
Col 3	668	PN92/944	*S. pneumoniae*	A	Eye
Col 5	669	PN93/832	*S. pneumoniae*		Blood
Col 6	670	PN93/779	*S. pneumoniae*	A	Blood
Col 7	671	PN93/356	*S. pneumoniae*	A	URT[1]
Col 8	672	PN93/707	*S. pneumoniae*	A	Sputum
Col 9	673	PN93/904	*S. pneumoniae*	A	Blood
Col 11	674	PN93/688	*S. pneumoniae*	A	CSF[2]
Col 12	675	PN93/950	*S. pneumoniae*	A	Blood
Col 13	676	na	na		na
Col 14	677	PN93/789	*S. pneumoniae*	A	Blood
Col 15	678	PN92/1139	*S. mitis*	B	Sputum
Col 16	679	PN93/952	*Streptococcus sp.*	B	LRT[3]
Col 17	680	PN91/2745	*Streptococcus sp.*		na
Col 18	681	PN93/454	*Streptococcus sp.*	B	Sputum
Col 19	682	PN93/800	*S. oralis*	B	Blood
Col 20	683	PN93/776	*Streptococcus sp.*	B	na
Col 21	684	PN93/447	*Streptococcus sp.*	C	Eye
Col 22	685	PN93/1264	*Streptococcus sp.*	B	na
Col 24	686	PN93/656	*Streptococcus sp.*		Sputum
Col 25	687	PN93/1003	*S. oralis*	C	Blood
Col 26	688	PN93/403	*Streptococcus sp.*		Synovial fluid
Col 27	689	PN93/135	*Streptococcus sp.*		Sputum
Col 28	690	PN93/918	*Streptococcus sp.*	B	Sputum

Table 2.6 – Additional streptococci used. * as defined in Whatmore *et al.* (Whatmore *et al.*, 2000), 1; URT –upper respiratory tract, 2; CSF – cerebrospinal fluid, 3; LRT – lower respiratory tract, na = data not available.

2.1.3 - *E. coli* Strains

Recombinant plasmids were maintained in *E. coli* JM109 (Yanisch-Perron *et al.*, 1985). This strain has the genotype *endA1, recA1, gyrA96, thi, hsdR17,* (r_k^-, m_k^+), *relA1, supE44, Δ(lac-proAB),* [F', *traD36, proAB, lacI*^{q}ZΔM15].

2.1.4 – Plasmids used

Molecular cloning was routinely carried out using either the pGEM-T™ TA cloning vector (Promega Corp.) or pUC18 (Yanisch-Perron *et al.*, 1985). Further plasmid constructs used in the generation of a Hyl$^-$:Crm^R mutant (see chapter 3, section 3.4.1) were derived from *hyl* maintained in pUC18.

Blue/white selection using the pGEM-T™ system was carried out according to the manufacturers recommendations.

During pneumococcal transformation experiments a control plasmid generated by Sue Colby was used. This plasmid (p683CAT) contains the pneumolysin gene interrupted by the chloramphenicol resistance cassette marker.

2.1.5 – Commonly used buffers and solutions

TE ; Tris-HCl [10mM], EDTA [1mM], pH 7.5-8.
5 X TAE; Tris-base [2M], Acetic acid [1M], EDTA [500mM].
5 X TBE; Tris-base [0.44M], Boric acid[0.44M], EDTA [10mM].
Stains-all solution; 'Stains-all' (1-ethyl-2-(3-[1-ethylnaptho(1,2-d)thiazolin-2-ylidine]-2-methylpropenyl)naptho-(1,2-d)thiazolium Bromide [0.1mM], Acetic acid [1mM], Ascorbic acid [0.5mM] in 50:50 (w/v) dH_2O 1,4-dioxan. Solutions were prepared prior to each assay, as prolonged storage was found to be detrimental to assay specificity.
Pneumococcal resuspension buffer; Tris-HCl [50mM], pH 7.5.
Pneumococcal hyaluronate digest buffer; Na_2HPO_4 [200mM], pH6.5, HA [2mgml^{-1}].
Ammonium acetate hyaluronate lyase assay buffer; Ammonium acetate [50mM], $CaCl_2$ [3mM], pH 6.5.
Gel loading buffer; Sucrose (50% w/v), EDTA [5mM], Bromophenol Blue (0.1% w/v), Xylene Cyanol FF (0.1% w/v), SDS (0.01% w/v).

2.1.6 - Kits routinely used

QIAPrep Spin Miniprep Kit (Cat. No. 27104) – Qiagen Ltd.
QIAQuick Gel Extraction Kit (Cat. No. 28704) - Qiagen Ltd.
QIAQuick PCR Purification Kit (Cat. No. 28106) - Qiagen Ltd.
pGEM®-T Easy Vector System II (Cat. No. A1380) – Promega Ltd.
Dig Hi-Prime (Cat. No. 1585606) – Boehringer-Mannheim Biochemica.
Dig Easy-Hyb (Cat. No. 1603558) – Boehringer-Mannheim Biochemica.

CEQ DNA Sequencing Kit (Cat. No. 60800) – Beckman Instruments Ltd.

2.1.7 - Primers used

Primer Name	Gene accession no. and base at 5' extremity of annealing site	Sequence (5'-3')	Annealing Temp (°C)
whp.for	L20670/20	GGCCCCGCGGAGTCATTGAGGCTAAGGAT	55
whp.rev	L206703004	GGCCCCGCGGTTGGCCTTCTTCGGATTTT	55
hyl.for	L20670/213	TGCAACGACGTCAGGAACAAAG	58
hyl.rev	L20670/2922	ACGGAATAAATAAAACGCCCCAAGTA	58
hspF1	L20670/285	AATCAAGCTGGAGTTATTC	55
hspF1B	L20670/623	GGATGACTGGAATGGCATCATCGC	55
hspF2	L20670/806	TGCCACTTATCGGAAATTG	55
hspF3	L20670/1078	CAGATCCCGAACATTTCCGAAAGA	55
hyl3.mid	L20670/1505	CGTGGCCGCAGTAGAAGTACTAAGA	58
hspF4	L20670/1877	ACTATAGCGATGGCTACTGGCCAA	55
hspF5	L20670/2232	GAAGCCTCCCTTACAGAACAAGAA	55
hspF6	L20670/2613	ATTCTGTCTCTACTATTTCCAACC	55
hspR1	L20670/2736	CTGGAGCTGATTCCTGGGT	55
hspR1b	L20670/2453	TTGCTTATGAGCCTGACTAATCGT	55
hspR2	L20670/2167	GCAGCAGTATCTGTTGAA	55
hyl5.mid	L20670/1811	CAAGGTACGACTGGAAAAGAGTGAC	58
hspR4	L20670/1199	AATTTCTTGATCATCCTTACGCAG	55
hspR5	L20670/961	CAGTTCCCAACAATGCTCTTTTC	55
hspR6	L20670/533	GGTTGTCCCTTCCTTAAGAGGTTC	55
hip5.out	L20670/158	CCCGACTTTATTATCTGT	50
hip3.out	L20670/3738	CGTGAAGAAATCATCCAC	52
hyp2.for	L20670/1270	GGATCCTATATCGACCACACCAATGTTG	48
hyp2.rev	L20670/2131	CCTAAAAAGGCAATCTTATC	48
gpx.for	Ctg4154*/13555	CCCTTGCAATCAGTTTATG	55
gpx.rev	Ctg4154*/13224	TTCAAAGACTTGCCCATCT	55
gpx5.out	Ctg4154*13496	CAGAAGGCGTTGATTTCC	52
cat.for	K01998/128	TAGTTCAACAAACGAAAATTGGATAA	53
cat.rev	K01998/1000	ACGGGGCAGGTTAGTGACAT	53
hexA.for	Ctg4241*/173	CCGACAATCCGATCCCTATGG	50
hexA.rev	Ctg4241*/2270	TATCAGCTGGTCCCGGTTCAATC	50
recP.for	Ctg4268*/228	ACCGCGACCGCTTTATTCTTTC	58
recP.rev	Ctg4269*/351	ATGCTGACTACGCGGGATTTTTC	58
TA.5	pGEM-T™**/158	TGACCATGATTACGCCAAGCT	58
TA.3	pGEM-T™**/2948	TCCCAGTCACGACGTTGTA	58
PE1	This Study/348	CCTTCTGGCTTTGCTGAATGGTAT	60

Table 2.7 – Primers used in this study
*Ctg denotes the contig upon which the gene lies at
http://wit.mcs.anl.gov/WIT2/CGI/org.cgi
** From the pGEM®-T Easy Vector System II - Promega Ltd.

2.1.8 - Other chemicals used

Sodium hyaluronate (from human umbilical chord) (Cat.No. 53740) – Fluka Chemicals.

CSP1 – Amino acid sequence N – EMRLSKFFRDFRQRKK – C

CSP2 – Amino acid sequence N – EMRISRIILDFLFLRKK – C

2.2 - Methods

2.2.1 – DNA manipulations

Pneumococcal genomic DNA extraction

Cells from 2X BHI/blood agar plates were harvested in a final volume of ≈ 500µl TE using a cotton swab. 10µl lysozyme [10 mg ml^{-1}] was added and suspensions were incubated at 37°C for 15-30 minutes. 30µl proteinase K [10 mg ml^{-1}] was added and suspensions were incubated at 37°C for a further 30 minutes. Following this, 40µl sodium sarkosyl (20%w/v) was added and suspensions were incubated at 37°C for 10 minutes. Lysates were mixed with 500µl phenol-chloroform (50:50 v/v) and spun at 13000 r.p.m. in a benchtop microcetrifuge. The supernatant was removed, and the remainder of the lysate was re-extracted with 200µl phenol-chloroform. The combined supernatants were again extracted with an equal volume of phenol-chloroform, then chloroform alone. DNA precipitation was performed by the addition of a 10% volume of sodium acetate ([3M], pH6.5) and 250% volume ice-cold ethanol. After incubation at –20°C for 15 minutes, the precipitate was collected by centrifugation at 13000 r.p.m. for 15 minutes. The pellet was washed with 70% ice-cold ethanol and centrifugation was repeated. The DNA pellet was dried for 40 minutes in a vacuum dessicator, and resuspended in 50-100µl dH_2O. This was followed by RNAse treatment at 65°C for 10 minutes. DNA was stored at –20°C.

PCR

PCR amplification was performed using *Taq* DNA polymerase supplied by Life Technologies Inc. according to the manufacturers instructions. Primer annealing temperatures were approximated using the Walace rule, where T_d (^{O}C) ≈ [2X(A+T)]+[4X(G+C)] (Wallace *et al.*, 1979).

Each reaction was conducted using $MgCl_2$ at 1.5μM for 35 cycles. Primer annealing temperatures are given in table 2.7. In reactions where two primers with different annealing temperatures were used, the lower temperature was included in the annealing step.

Inverse PCR (iPCR)

This technique is used to identify sequences 5' or 3' to a given, known DNA sequence.

Restriction digests are selected which do not cut, or cut in a defined portion of the known sequence. Endonuclease treatment was carried out for 2 hours at the recommended temperature, then heat inactivated by incubation at 80^{O}C for 10 minutes. Digest fragments are then treated with T4 DNA ligase for 4 hours at room temperature. Theoretically, a circularised product which contains the known DNA sequence, plus an indeterminate portion of the flanking sequence, defined by the position at which the chosen enzyme cuts outside the known sequence. The outward facing primers are then used to amplify this region from the lariat molecule formed in the ligation. This approach is outlined diagrammatically in figure 2.1, below.

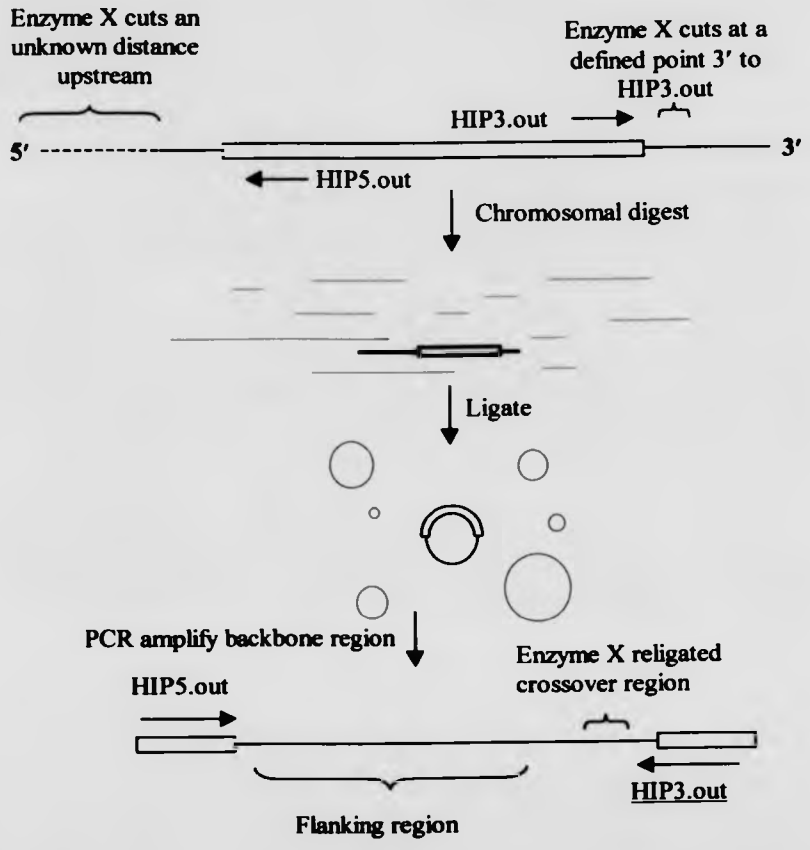


Figure 2.1 – Diagrammatic representation of the iPCR technique.

Restriction endonuclease treatment

Restriction digests were performed according to the suppliers recommendations. During the HRRA experiments it was found that the addition of BSA to a final concentration of 100μgml^{-1} improved the quality of digest patterns after PAGE.

DNA electrophoresis

DNA was routinely visualised by electrophoresis through either 0.8-1% agarose in TAE buffer, or 4% and 8% polyacrylamide in TBE buffer. Polyacrylamide gels were

stained post-electrophoresis in 1XTBE containing ethidium bromide at $60\mu gl^{-1}$. Agarose gels were cast containing ethidium bromide at $35\mu gl^{-1}$, and the running buffer contained ethidium bromide at $60\mu gl^{-1}$. Stained DNA was visualised on an ultraviolet light source at 300nm (UVP Inc.).

High Resolution Restriction Analysis (HRRA)

In previous studies HRRA has been used to estimate the diversity between alleles of a particular gene from large numbers of strains. This is followed by a statistical analysis of the data generated. (Muller-Graf *et al.*, 1999, Whatmore and Dowson, 1999, Doherty *et al.*, 2000). The method is illustrated diagrammatically in figure 2.2.

Briefly, aliquots of a given PCR product are independently subjected to restriction digestion with a range of frequently cutting enzymes (normally with 4 base specificities). Products are separated by PAGE through 4% and 8% polyacrylamide gels. Differences in the nucleotide sequences of distinct alleles results in the loss or gain of restriction sites, and hence alters the band pattern of a given allele when digested with a given enzyme. Different band patterns arising from digestion with a common enzyme are referred to as restriction types. The restriction types are combined to give alleles. Alleles are differentiated on the basis of different combined restriction types. Comparisons are then made between strains by looking at the distribution of common and divergent bands against the total number of digest products in all restriction types. Comparisons were made by eye and variations in electrophoretic mobility were controlled by the inclusion of multiple reference lanes containing DNA mass ladder. Poor staining and visualisation of smaller digest fragments complicated the analysis. As such, fragments of <50bp were not included in the band matching procedure.

Statistical analysis of the results was conducted using the algorithm of Nei and Li (Nei and Li, 1979).

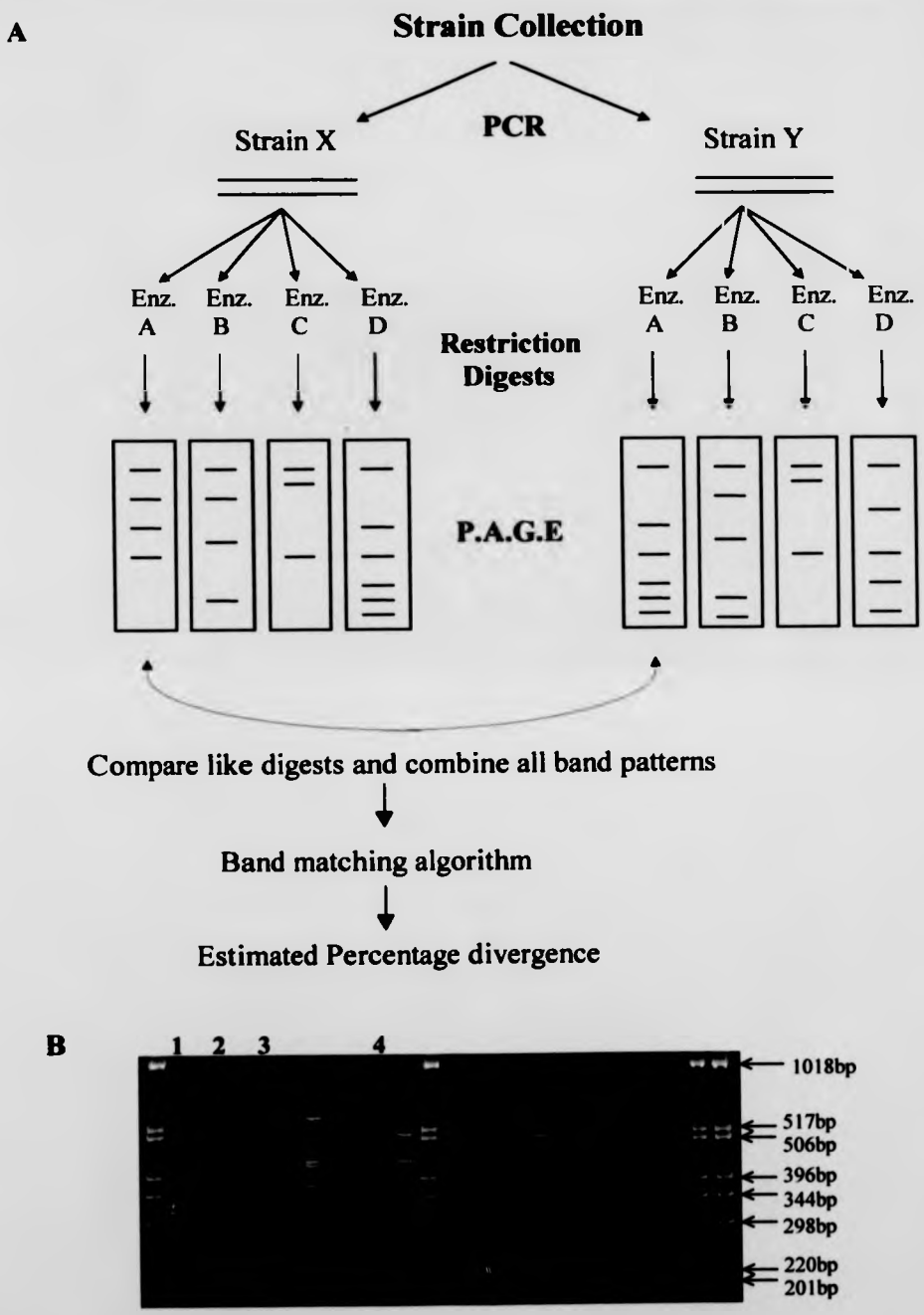


Figure 2.2 – A; Diagrammatic representation of the HRRA technique. B; Example of a typical HRRA PAGE image (8% acrylamide gel). Individual restriction types are numbered.

DNA ligations

Ligations were performed using T4 DNA ligase supplied by Life Technologies Inc. according to the manufacturers recommendations.

Construction of T-vectors

5μg pUC18 was digested with a blunt cutting restriction endonuclease (*Bst*1107I), which cleaves pUC18 once. The linearised vector was electrophoresed through agarose and recovered using the QIAgen Gel Extraction Kit. 5.5μl 10X *Taq* buffer was added to the eluate volume of 50μl, and 2μl dTTP [100mM]. 10 units of *Taq* polymerase was added to the mixture, which was mixed and incubated at 72^{O}C for 2 hours. Once again, the linearised-tagged vector was purified by agarose gel electrophoresis and recovered with the QIAgen kit. Untagged vector was re-circularised with 2 units T4 DNA ligase, in the appropriate ligase buffer. The entire volume was electrophoresed once more, and the linear plasmid excised and purified. Ligations were performed immediately at vector:insert ratios ranging from 0.1:1 to 1:0.1.

Southern blotting

Southern analysis was performed according to the directions provided by Boehringer-Mannheim Biochemica in their guide to Southern blotting. Amendments made due to the use of alternative reagents, such as EasyHyb™ are detailed in the instructions provided with the reagents.

DNA Sequencing

Two methods of DNA sequencing were used during the course of the work described here; the CEQ2000 DNA Analysis System – Beckman Instruments Inc. and the ABI Prism 377 DNA Sequencing System – Applied Biosystems. Reactions were carried out using materials and procedures provided by the manufacturers.

Sequence analysis was performed using the DNAStar suite of programs (DNAStar Inc.), and programs within the MEGA suite of programs (Kumar *et al.*, 1993).

2.2.2 - Bacterial transformation

E.coli

Competent *E.coli* (JM109) were prepared by the $CaCl_2$ method described in Sambrook *et al.* (Sambrook *et al.*, 1989). The heat shock method of transformation included within the same reference was also used.

S. pneumoniae

Section 1.3 of chapter 1 describes the growth-phase dependent nature of pneumococcal competence for genetic transformation. In our lab it is common practice to define a growth/competence curve for each given strain. This involves C-medium adapting cells prior to starting the experiment. *S. pneumoniae* were harvested in 5ml C-medium after overnight growth on BHI/blood agar. This was incubated at 37^{O}C for 2-4 hours, or until the suspension acquires a slightly turbid appearance. Aliquots were then frozen at –80^{O}C after the addition of 15% (v/v) glycerol. These were used as starting inocula in experiments to define the point at which the cultures became competent.

100µl of C-medium adapted cells were inoculated in 7ml C-medium and incubated at 37^{O}C until the OD_{550} reached 0.2 from this point onwards the culture was sampled every 15 minutes and frozen in 15% (v/v) glycerol.

p683CAT was used as a positive control plasmid to test the aliquots from each time point for competence. This was carried out by transformation as follows; 25µl frozen stocks were added to 500µl C-medium. 8µl BSA (10% w/v), 500ng test plasmid, 0.5µl $CaCl_2$ [1M] and 2µl of either CSP1 or CSP2 was then added. This was incubated at 37^{O}C for 3 hours, and then spread on selective BHI/blood agar. Positive transformants were identified after incubation at 37^{O}C for 12-20 hours.

2.2.3 – HA digests and hyaluronate lyase assays

Hyaluronate digests

The method of hyaluronate digests is taken from the method of Homer *et al.* (Homer *et al.*, 1996, and *personal communication*). Briefly, cells were harvested from an overnight culture of *S. pneumoniae* grown on BHI/blood agar in 500µl pneumococcal resuspension buffer (see section 2.1.5). The OD_{620} of the suspension was adjusted to 0.1 (+/- 15%) when read in an iEMS microtitre plate reader (Labsystems). This crude enzyme preparation was then used in hyaluronate digests, which were prepared as follows; 100µl cell suspension was added to 350µl pneumococcal hyaluronate digest buffer (section 2.1.5) and 200µl aqueous HA [2mg/ml]. Digestion was allowed to proceed for 24 hours at 37°C. Negative controls were performed using resuspension buffer as an inoculum.

Induction with hyaluronate

Induction of hyaluronate lyase was achieved by incubation of cells harvested from overnight growth on solid media in 1ml C-medium (made without glucose or sucrose) to which HA had been added to a final concentration of $0.5mgml^{-1}$. This was incubated at 37°C for 3 hours. Inclusion of glucose in this media had no effect on induction phenotype. Induction mixtures were monitored at OD_{620}, which indicated that growth was not observed during this time.

At the end of this incubation, cells were harvested by cenrifugation at 3000 r.p.m. for 5 minutes. The cell pellet was washed once with pneumococcal resuspension buffer, and resuspended in 500µl of the same. As for non-induced cells, the $O.D_{620}$ was adjusted to 0.1 (+/- 15%), and the digests performed as described above. Induction was defined as an increase in titre by ≥20% using either assay.

'Stains-all' assay

5µl of the digest mix was added to 45µl 'Stains-all' solution [0.1mM]. 50µl dH_2O was added and A_{650} was recorded using the iEMS microtitre plate reader (Labsystems). Assays were performed in quadruplicate, with a standard HA curve defined during each set of assays (over the range 0.5-0.05 $mgml^{-1}$).

The standard deviation and 95% confidence limits for each set of quadruplicate results were defined as shown in the equation below. Mean values were compared to the mean value obtained for the digest negative control. One unit of activity was defined as a decrease in 10% of the assay sample as compared to the negative control value, after Benchetrit *et al.* (Benchetrit *et al.*, 1977).

$$95\%\text{CL for } \mu = x \pm \frac{ts}{\sqrt{N}} \text{ where } s = \sqrt{\frac{\sum y^2}{(n-1)}}$$

Where μ = population mean
x = sample mean
n = sample size
t = t statistic for 95% confidence internal at N = 3
s = standard deviation of each quadruplicate sample
N = degrees of freedom (= sample size –1)

Unsaturated disacharride assay

This was adapted from the method of Lin, Averett and Pritchard (Lin *et al.*, 1997). Briefly, 30μl of digest mix was added to 60μl ammonium acetate hyaluronate lyase assay buffer. The absorbance was recorded at 232nm using a 0.1ml quartz cuvette, path length 0.5ml. The amount of 2-acetamido-2-deoxy-3-O-(β-D-gluco-4-enepyranosyluronic acid released was quantified according to the Beer-Lambert equation with the appropriate corrections for dilution and path length. The millimolar extinction coefficient (ε) used was 5.5 (Yamagata *et al.*, 1968).

2.2.4 – Statistical treatment of hyaluronate lyase activity data

Mean values of hyaluronidase activity as measures by the 'Stains-all' assay, and released digest product were calculated for the strains within the groups tested. Variation of hyaluronate lyase activity between the groups was analysed using the analysis of variance, or Anova technique.

This treatment provides an output based on the ratio of inter- and intra-group variance, termed the F value (F_s). This statistic is used in the F-test in order to prove or disprove the null-hypothesis. The initial calculations of are outlined below, followed by a description of the F-test.

Each data set was treated in the following manner (x or x denotes value or reading, y or y denotes variance from mean);

Blood	Carried	Lung	CSF	Ear
x_{b1}	x_{c1}	X_{l1}	x_{csf1}	x_{e1}
x_{b2}	x_{c2}	X_{l2}	x_{csf2}	x_{e2}
x_{b3}	x_{c3}	X_{l3}	x_{csf3}	x_{e3}
x_{bn}	x_{cn}	X_{ln}	x_{csfn}	x_{en}
$\bar{x}_b$	$\bar{x}_c$	$\bar{x}_l$	$\bar{x}_{csf}$	$\bar{x}_e$
n_b	n_c	n_l	n_{csf}	n_e
$\sum X^{1-n}{}_b$	$\sum X^{1-n}{}_c$	$\sum X^{1-n}{}_l$	$\sum X^{1-n}{}_{csf}$	$\sum X^{1-n}{}_e$
$\sum y_b^{\,2}$	$\sum y_c^{\,2}$	$\sum y_l^{\,2}$	$\sum y_{csf}^{\,2}$	$\sum y_e^{\,2}$

Where $\bar{x}_i = \dfrac{\sum X^{1-n}{}_i}{n_i}$

n_i = Group sample size (blood = 20, carried = 49 etc).

$\sum X'_{1-n}$ = Sum of values for that group sample.

$$\sum y_i^{\,2} = \left|(x_{i1} - \bar{x}_i)^2 + (x_{i2} - \bar{x}_i)^2 + (x_{i3} - \bar{x}_i)^2 \ldots\ldots (x_{in} - \bar{x}_i)^2\right|$$

Subscripts denote the group of origin, b = blood, c = carried etc.

1. The sample mean ($\bar{\bar{x}}_{b-e}$) across the groups is calculated;

$$\bar{\bar{x}}_{b-e} = \frac{1}{\sum n_i} \times \sum X_{b-e}$$

2. The inter-group variance ($\bar{Y}_1$, weighted by group size, n_i) is given by;

$$\bar{Y}_1 = \sum^{b-e} n_i (\bar{x}_i - \bar{\bar{x}}_{b-e})^2$$

3. The intra-group variance ($\overline{Y}_2$) is given as the sum of variances within each group;

$$\overline{Y}_2 = \sum^{b-e}\sum y_i^2$$

4. The total sum of the squares (SS_T, intra+inter) is given by;

$$SS_T = \sum^{b-e} n_i\left(\bar{x}_i - \bar{\bar{x}}_{b-e}\right)^2 + \sum^{b-e}\sum y_i^2$$

The next stage of the anova is to tabulate these values as follows;

Source of Variation	DF	SS	MS	f_s
$\bar{x}_i - \bar{\bar{x}}_{b-e}$ (Inter)	$a-1$	$\overline{Y}_1$	$\frac{\overline{Y}_1}{a-1} = F_1$	$\frac{F_1}{F_2}$
$x_i - \bar{x}_i$ (intra)	$\left(\sum^{b-e} n_i\right) - a$	$\overline{Y}_2$	$\frac{\overline{Y}_2}{\sum n_{b-e} - a} = F_2$	
$x_i - \bar{\bar{x}}_{b-e}$ (total)	$\sum^{b-e} n - 1$	$\overline{Y}_1 + \overline{Y}_2$		

Where DF = Degrees of Freedom

SS = Sum of Squares

MS = Mean of Squares

a = Number of groups (in this case a=5)

At this stage it is possible to carry out the final F-test. The basis of the F-test is the comparison of the F_s output of the anova with a previously determined value, F_{crit}. This latter value is based on a theoretical probability distribution whose derivation is beyond the scope of the work described here. F_{crit} can be defined as the value below which the null-hypothesis is obeyed at a given level of confidence. When $F_s<F_{crit}$, the null-hypothesis is deemed true, and the groups are said not to vary significantly from one another. When $F_s>F_{crit}$ then the null-hypothesis is disproved, indicating that there is an added element of variation, ie; a true difference between the groups studied. In order to determine F_{crit}, certain experimental parameters must be defined. Firstly, the experimental degrees of freedom are considered (outlined in the anova table above). Secondly, the expected level of confidence (α) is set. It is standard

practice to take α=0.05. Taking these values, and consulting tables of the F-distribution enables the user to define the F_{crit} values for those given parameters, that is, the F value below which the null-hypothesis is obeyed at a given level of confidence. The majority of the calculations in the work described here were performed using the readily available Microsoft Excel data analysis add-in, which precisely calculates F_{crit}, and also gives a probability value.

2.2.4.1 - Other statistical methods

The Tukey-Kramer method was used to perform post-hoc analysis of the Anova results. This test enables us to make pairwise comparisons of group means, and asks the question 'Is the variation seen higher than that expected from a predetermined value?'. This value is derived from the studentised Q distribution, and is called the Most Significant Difference' of MSD. It is calculated as follows;

$$\text{MSD} = Q_{[K,V]} \text{ X Standard Error}_{[A,B]}$$

$$\text{Where Standard Error}_{[A,B]} = \sqrt{\frac{MS(within)X\left(\frac{1}{nA}\right)+\left(\frac{1}{nB}\right)}{2}}$$

$Q_{[K,V]}$ is derived from statistical distribution of the Q value for degrees of freedom [K,V], where K = number of groups used and V = Degrees of freedom of MS(within). Suffixes A and B denote individual groups (e.g.; Carried and Blood isolates). When the observed value of the differences between the means of two individual groups exceeds the value derived for MSD, the difference is said to be significant.

Various statistical analyses were performed using the chi-squared (χ^2) test. This test is based on a comparison of the observed and expected frequencies of a particular value. This is compared against the χ^2 distribution to determine whether the difference of the distribution of values is significant. The value for χ^2 is determined according to the following equation;

$$\chi^2 = \sum \frac{(O-E)^2}{O}$$

Where O = observed value
E = expected value

High values for χ^2 indicate that a particular distribution has a low probability of occurring by chance.

The Maximum χ^2 test of Maynard-Smith was used in the detection of recombination between allelic variants (Maynard-Smith, 1992). This technique compares the distribution of polymorphisms along donor and recipient sequences with those expected to occur by chance. The test looks at every position along aligned sequences and generates a contingency matrix of the number of polymorphisms either side of a given position. The value of χ^2 is determined for this matrix, and the position at which the maximum value of χ^2 is obtained is identified as the putative recombination point.

The significance of this result is tested by running trials where random sequences of the same length (and polymorphic sites) are created. The significance level is determined as the proportion of trial pairs with greater maximum χ^2 values than the observed data.

2.2.5 – Biofilm growth model

Biofilms were generated using the Sorbarod filter model of Gilbert *et al.* (Gilbert *et al.*, 1989, Hodgson *et al.*, 1995). This model establishes a biofilm culture on the parallel cellulose acetate fibres of the Sorbarod filter. A diagram of the Sorbarod biofilm system is shown in figure 2.3.

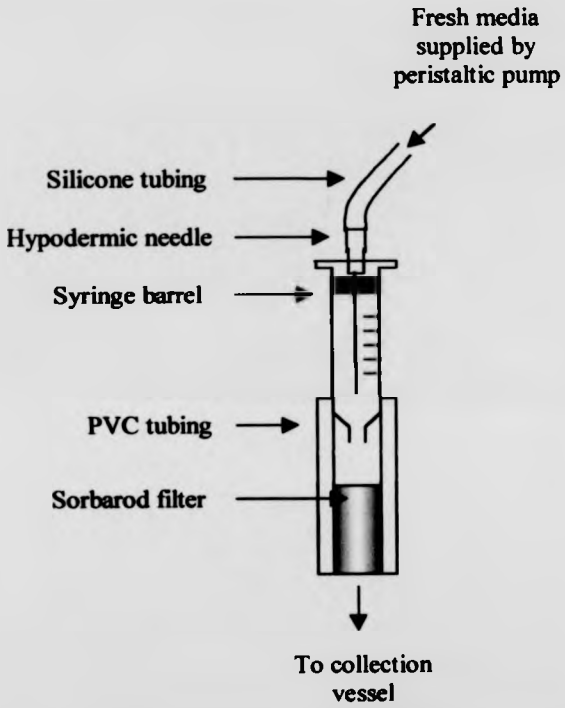


Figure 2.3 – Diagrammatic representation of the Sorbarod model

The experiment was set up by inoculating the Sorbarod filter with the planktonic culture of choice, and supplying fresh liquid media at a constant rate. As the biofilm became established, acapsular variants arose. This was shown by either sacrificing the filter at a given time point and culturing the organisms on solid media, or by culturing the cells which were shed in the effluent.

3.1 Occurrence and variation of the *hyl* locus

3.1.1.1 - Introduction

Previous studies have addressed the question of genetic diversity at loci encoding virulence determinants of pathogenic streptococci. This has shown that in some instances variation can be high, and that the evolution of these virulence determinants is in some cases mediated by homologous recombination between related genes (Poulsen *et al.*, 1998, Whatmore and Dowson, 1999, S. King, unpublished data).

The first question addressed in this study was that of genetic diversity at the *hyl* locus. In previous studies this has been achieved using high resolution restriction analysis (HRRA), also known as restriction fragment length polymorphism (RFLP) (Muller-Graf *et al.*, 1999, Whatmore and Dowson, 1999, Doherty *et al.*, 2000), followed by a statistical analysis of the data generated. This is a technique suited to high throughput analysis of loci from large numbers of strains, and is a cost effective alternative to direct sequencing of loci from all isolates in the first instance. The technique is described in chapter 2. The band comparisons are quantified using the band matching algorithm described by Nei and Li (Nei and Li, 1979). The output of this procedure is presented as percentage divergence between alleles found.

Table 2.1 shows the strains used in this study, which includes 52 strains isolated from asymptomatic carriers, 54 invasive disease isolated and 35 serotype 3 strains, a serotype often associated with invasive disease (Gray, 2000).

In this study two other loci were studied by HRRA. These were *recP* (Accession No. M31296) and *hexA* (Accession No. M18729), encoding proteins involved in recombination and mismatch repair respectively (see Chapter 1, section 1.3). This was in order to enable a direct comparison between allelic variation in *hyl*, encoding a virulence determinant, and two loci encoding genes with housekeeping functions.

The *hyl* fragment used in this study was amplified using the hyl.for/rev primer pair, described in Chapter 2, section 2.1.7. The predicted size of this

product is 2695 bp. Amplification was successful from 133/141 strains (94.3%). PCR failure was assumed to be as a result of polymorphism at one or both of the primer binding sites on the target sequence. The presence of *hyl* was confirmed by Southern analysis (see section 3.1.2). The *recP* amplicon is calculated to be 1568 bp in length, and was successfully amplified from 141/141 (100%) of the strains. The *hexA* amplicon is 2097 bp long, and was amplified from 138/141 (98.5%) of the strains. This ratio of PCR success was taken as indicative of higher variation at the *hyl* locus over those of *hexA* or *recP*. This was found to be statistically significant by χ^2 analysis, with $\chi^2 = 9.416$, $P \leq 0.025$ (where $\chi^2_{crit} = 3.84$).

A series of trial digests was performed on the *hyl*, *hexA* and *recP* PCR products, and the enzymes finally chosen on the basis of the most clearly discernible digest patterns with the strains used. These are listed in table 3.1.1.

recP	*hexA*	*hyl*
*Hinf*I	*Alu*I	*Dde*I
*Hsp*92II	*Hha*I	*Hae*III
*Mae*III	*Hinf*I	*Hinf*I
*Mnl*I	*Hsp*92II	*Mse*I
*Tsp*509I	*Tsp*509I	*Rsa*I

Table 3.1.1 - Restriction enzymes used in HRRA of the loci used in this study.

The overall restriction endonuclease footprints of the database gene sequences of each locus were 8.4% for *recP*, 8.5% for *hexA*, and 7.71% for *hyl*.

3.1.1.2 - HRRA of *recP*

The restriction patterns seen with the *recP* digests were assigned restriction types, which were compiled as ET's. These were deemed synonymous with alleles. These data are summarised in table 3.1.12 in appendix 3.1.

The estimated percent divergence between each allele was calculated as follows: The total number of bands for each allele type were established by adding together the number of bands from each RT. Pairwise comparisons between each RT for all alleles (1 versus2, 1 versus 3, 1 versus 4 and so on) were made.

These data were tabulated, and are presented in table 3.1.2. The allele numbers are shown in bold, with the total number of bands for that allele displayed either above (on the upper X axis), or to the left (on the Y axis) of the allele number. The upper right triangular matrix gives the number of common bands between alleles. These data were used to compute the estimated percent divergence data using the Nei 3 program (Nei and Li, 1979). The individual pairwise values of this calculation are shown in the lower left triangular matrix.

Original
Contains
Pullouts

pgs 75, 77 & 79

recP	tot.band	39	39	39	38	37	38	38	38	37	39	40	
tot. band	allele	1	2	3	4	5	6	7	8	9	10	11	
39	1	X	34	32	36	35	36	34	34	35	35	34	
39	2	1.156	X	30	33	32	33	34	31	32	32	31	
39	3	1.675	2.232	X	34	30	33	30	35	34	35	36	
38	4	0.562	1.3	1.046	X	34	35	33	33	34	37	37	
37	5	0.689	1.452	2.007	0.823	X	34	32	32	33	34	32	
38	6	0.562	1.3	1.3	0.689	0.823	X	35	36	34	35	36	
38	7	1.046	1.046	2.12	1.189	1.338	0.689	X	33	34	34	33	
38	8	1.046	1.837	0.8	1.189	1.338	0.452	1.189	X	34	35	34	
37	9	0.689	1.452	0.935	0.823	0.962	0.823	0.823	0.823	X	33	32	
39	10	0.909	1.675	0.909	0.332	0.935	0.8	1.046	0.8	1.189	X	35	
40	11	1.264	2.058	0.779	0.44	1.564	0.671	1.411	1.156	1.564	1.017	X	
38	12	0.8	1.046	2.12	0.935	0.578	0.935	1.452	1.452	1.076	1.3	1.675	
38	13	1.046	1.837	0.562	0.452	1.338	2.007	1.724	0.689	0.341	0.8	1.156	1.
41	14	1.893	2.433	0.429	1.264	2.232	1.52	2.343	1.264	1.156	1.124	0.991	2.
37	15	1.724	2.607	1.452	1.893	2.07	1.076	1.893	0.578	1.495	1.452	1.837	2.
38	16	1.837	1.3	1.3	0.935	0.823	0.689	1.189	1.189	0.823	1.564	0.909	0.
36	17	0.823	1.338	1.61	0.962	1.108	0.465	0.228	0.962	1.108	1.338	1.189	1.

Table 3.1.2 - Band matching data and estimated percentage divergence of *recP*.

37	38	38	38	37	39	40	38	38	41	37	38	36
5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
35	36	34	34	35	35	34	35	34	32	31	31	34
32	33	34	31	32	32	31	34	31	30	28	33	32
30	33	30	35	34	35	36	30	36	38	32	33	31
34	35	33	33	34	37	37	34	36	34	30	34	33
X	34	32	32	33	34	32	35	32	30	29	34	32
0.823	X	35	36	34	35	36	34	30	33	33	35	35
1.338	0.689	X	33	34	34	33	32	31	30	30	33	36
1.338	0.452	1.189	X	34	35	34	32	35	34	35	33	33
0.962	0.823	0.823	0.823	X	33	32	33	36	34	31	34	32
0.935	0.8	1.046	0.8	1.189	X	35	33	35	35	32	32	32
1.564	0.671	1.411	1.156	1.564	1.017	X	32	34	36	31	35	33
0.578	0.935	1.452	1.452	1.076	1.3	1.675	X	32	30	29	34	32
1.338	2.007	1.724	0.689	0.341	0.8	1.156	1.452	X	35	32	33	31
2.232	1.52	2.343	1.264	1.156	1.124	0.991	2.343	1.017	X	31	33	30
2.07	1.076	1.893	0.578	1.495	1.452	1.837	2.186	1.338	1.948	X	30	30
0.823	0.689	1.189	1.189	0.823	1.564	0.909	0.935	1.189	1.52	1.893	X	33
1.108	0.465	0.228	0.962	1.108	1.338	1.189	1.224	1.495	2.12	1.66	0.962	X

3.1.1.3 - HRRA of *hexA*

The different *hexA* alleles were assigned on the basis of overall restriction and ET's, allele assignments are shown in table 3.1.13 in appendix 3.1.

These data were treated in the same way as those obtained in the *recP* study, and the data matrix is shown in table 3.1.3.

hexA	Tot. Bands	46	43	46	45	44	43	45	44	45	44	46
Tot. Bands	Allele	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
46	*1*	X	41	43	44	42	40	43	42	43	37	42
43	*2*	0.687	X	41	42	40	37	41	40	38	40	40
46	*3*	0.565	0.687	X	44	43	41	43	42	43	38	42
45	*4*	0.28	0.389	0.28	X	43	40	44	43	44	38	43
44	*5*	0.578	0.703	0.38	0.286	X	39	42	41	42	38	41
43	*6*	0.896	1.267	0.687	0.8	0.918	X	39	38	39	34	38
45	*7*	0.473	0.591	0.473	0.187	0.484	1.015	X	42	43	39	44
44	*8*	0.578	0.703	0.578	0.286	0.591	1.138	0.484	X	42	36	41
45	*9*	0.473	1.236	0.473	0.187	0.484	1.015	0.38	0.484	X	37	42
44	*10*	1.657	0.703	1.428	1.332	1.236	2.094	1.111	1.699	1.561	X	38
46	*11*	0.763	0.896	0.763	0.473	0.781	1.332	0.28	0.781	0.671	1.428	X
41	*12*	2.094	2.592	2.353	1.994	2.418	2.049	2.253	2.418	1.994	3.916	2.621
43	*13*	0.484	0.82	0.687	0.389	0.703	0.398	0.591	0.703	0.591	1.843	0.896
43	*14*	0.687	0.82	0.687	0.389	0.703	1.267	0.591	0.703	0.591	1.366	0.896
44	*15*	0.38	0.495	0.578	0.094	0.591	0.918	0.484	0.389	0.286	1.464	1.428

Table 3.1.3 - Band matching data and estimated percentage divergence of *hexA*.

	45	44	43	45	44	45	44	46	41	43	43	44
	4	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***
	44	42	40	43	42	43	37	42	34	42	41	43
	42	40	37	41	40	38	40	40	31	39	39	41
	44	43	41	43	42	43	38	42	33	41	41	42
3	X	43	40	44	43	44	38	43	34	42	42	44
3	0.286	X	39	42	41	42	38	41	32	40	40	41
7	0.8	0.918	X	39	38	39	34	38	33	41	37	39
3	0.187	0.484	1.015	X	42	43	39	44	33	41	41	42
3	0.286	0.591	1.138	0.484	X	42	36	41	32	40	40	42
3	0.187	0.484	1.015	0.38	0.484	X	37	42	34	41	41	43
3	1.332	1.236	2.094	1.111	1.699	1.561	X	38	27	35	37	37
3	0.473	0.781	1.332	0.28	0.781	0.671	1.428	X	32	40	40	38
3	1.994	2.418	2.049	2.253	2.418	1.994	3.916	2.621	X	35	31	33
7	0.389	0.703	0.398	0.591	0.703	0.591	1.843	0.896	1.541	X	39	41
7	0.389	0.703	1.267	0.591	0.703	0.591	1.366	0.896	2.592	0.82	X	40
3	0.094	0.591	0.918	0.484	0.389	0.286	1.464	1.428	2.151	0.495	0.703	X

3.1.1.4 - HRRA of *hyl*

The different *hyl* alleles were assigned on the distribution of restriction types and are summarised in table 3.1.14 in appendix 3.1.

Once again, these data were treated in the same way as those obtained in the *recP* study, and the data matrix is shown in table 3.1.4.

hyl	Tot. Bands	49	49	49	50	50	49	46	50	49	49	48
Tot. Bands	Allele	1	2	3	4	5	6	7	8	9	10	11
49	1	X	45	41	46	48	46	41	47	48	46	44
49	2	0.714	X	45	46	47	48	43	48	46	45	44
49	3	1.507	0.714	X	42	43	44	40	44	42	41	43
50	4	0.614	0.614	1.387	X	48	47	41	47	45	44	48
50	5	0.257	0.433	1.186	0.341	X	48	42	49	47	45	46
49	6	0.529	0.172	0.904	0.433	0.257	X	43	47	45	46	45
46	7	1.241	0.836	1.452	1.33	1.124	0.836	X	43	42	41	39
50	8	0.433	0.257	0.991	0.518	0.168	0.433	0.924	X	48	44	45
49	9	0.172	0.529	1.3	0.8	0.341	0.714	1.035	0.257	X	45	43
49	10	0.529	0.714	1.507	0.991	0.8	0.529	1.241	0.991	0.714	X	42
48	11	0.817	0.817	1.012	0.172	0.529	0.628	1.578	0.714	1.012	1.213	X
50	12	0.257	0.8	1.594	0.699	0.341	0.614	1.541	0.518	0.8	1.186	0.904
48	13	1.419	0.817	0.817	1.1	1.1	1.012	1.578	0.904	1.213	1.63	0.924
47	14	1.33	1.124	1.983	1.419	1.213	1.33	0.462	1.012	1.124	1.759	1.669
48	15	1.213	0.262	0.442	0.964	0.714	0.442	0.945	0.529	0.817	1.012	1.124
47	16	0.924	0.924	1.759	1.213	0.817	1.124	0.273	0.628	0.73	1.541	1.452
49	17	0.714	0.529	1.3	0.8	0.433	0.714	1.241	0.257	0.529	1.3	1.012
48	18	0.442	0.442	1.012	0.714	0.529	0.262	0.945	0.714	0.442	0.262	0.924
41	19	2.388	2.388	2.649	2.484	2.232	2.136	2.094	2.484	2.649	1.893	2.82
49	20	1.3	1.1	0.172	1.806	1.594	1.3	1.669	1.387	1.1	1.3	1.419
49	21	0.529	0.714	1.507	0.614	0.257	0.529	1.452	0.433	0.714	1.1	0.817
47	22	2.214	1.759	1.541	1.848	1.848	1.541	2.674	2.073	2.453	2.214	1.893
50	23	0.8	0.433	1.186	0.699	0.518	0.614	1.124	0.341	0.614	1.186	0.904
48	24	0.442	0.628	0.817	0.529	0.172	0.442	1.361	0.348	0.628	1.012	0.355
47	25	0.73	0.355	1.541	1.012	1.012	0.924	1.27	0.817	0.54	0.924	1.241
48	26	0.817	0.628	1.419	0.348	0.529	0.442	1.361	0.714	1.012	1.012	0.54
48	27	0.817	0.442	1.213	0.714	0.529	0.262	1.15	0.714	1.012	0.817	0.924
49	28	0.529	0.714	1.507	0.614	0.257	0.529	1.452	0.433	0.714	1.1	0.817

Table 3.1.4 - Band matching data and estimated percent divergence of *hyl*.

50	50	49	46	50	49	49	48	50	48	47	48	47	49	48	41
4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
46	48	46	41	47	48	46	44	48	41	41	42	43	45	46	34
46	47	48	43	48	46	45	44	45	44	42	47	43	46	46	34
42	43	44	40	44	42	41	43	41	44	38	46	39	42	43	33
X	48	47	41	47	45	44	48	46	43	41	44	42	45	45	34
0.341	X	48	42	49	47	45	46	48	43	42	45	44	47	46	35
0.433	0.257	X	43	47	45	46	45	46	43	41	46	42	45	47	35
1.33	1.124	0.836	X	43	42	41	39	40	39	44	42	45	41	42	34
0.518	0.168	0.433	0.924	X	48	44	45	47	44	43	46	45	48	45	34
0.8	0.341	0.714	1.035	0.257	X	45	43	45	42	42	44	44	46	46	33
0.991	0.8	0.529	1.241	0.991	0.714	X	42	43	40	39	43	40	42	47	36
0.172	0.529	0.628	1.578	0.714	1.012	1.213	X	44	43	39	42	40	43	43	32
0.699	0.341	0.614	1.541	0.518	0.8	1.186	0.904	X	42	41	46	42	46	44	34
1.1	1.1	1.012	1.578	0.904	1.213	1.63	0.924	1.3	X	38	42	39	45	41	30
1.419	1.213	1.33	0.462	1.012	1.124	1.759	1.669	1.419	1.893	X	40	45	42	40	30
0.964	0.714	0.442	0.945	0.529	0.817	1.012	1.124	0.529	1.124	1.452	X	41	44	44	33
1.213	0.817	1.124	0.273	0.628	0.73	1.541	1.452	1.213	1.669	0.363	1.241	X	43	41	31
0.8	0.433	0.714	1.241	0.257	0.529	1.3	1.012	0.614	0.628	1.124	0.817	0.924	X	43	33
0.714	0.529	0.262	0.945	0.714	0.442	0.262	0.924	0.904	1.33	1.452	0.73	1.241	1.012	X	34
2.484	2.232	2.136	2.094	2.484	2.649	1.893	2.82	2.484	3.389	3.289	2.551	3	2.649	2.291	X
1.806	1.594	1.3	1.669	1.387	1.1	1.3	1.419	2.025	1.213	2.214	0.817	1.983	1.719	1.012	2.91
0.614	0.257	0.529	1.452	0.433	0.714	1.1	0.817	0.614	1.419	1.541	1.012	1.124	0.714	0.817	2.38
1.346	1.848	1.541	2.674	2.073	2.453	2.214	1.893	2.073	2.361	3.301	1.669	3.03	2.453	1.893	3.28
0.699	0.518	0.614	1.124	0.341	0.614	1.186	0.904	0.699	1.1	0.628	0.714	1.012	0.433	0.904	2.74
0.529	0.172	0.442	1.361	0.348	0.628	1.012	0.355	0.529	0.924	1.452	0.924	1.035	0.628	0.73	2.55
1.012	1.012	0.924	1.27	0.817	0.54	0.924	1.241	1.213	0.641	1.578	1.035	1.361	0.54	0.641	3
0.348	0.529	0.442	1.361	0.714	1.012	1.012	0.54	0.714	1.124	1.452	0.924	1.452	1.012	0.73	2.55
0.714	0.529	0.262	1.15	0.714	1.012	0.817	0.924	0.904	1.33	1.669	0.73	1.452	1.012	0.54	2.55
0.614	0.257	0.529	1.452	0.433	0.714	1.1	0.817	0.433	0.817	1.33	1.012	1.124	0.172	0.817	2.38

48	47	48	47	49	48	41	49	49	47	50	48	47	48	48	49
13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
41	41	42	43	45	46	34	42	46	37	45	46	44	44	44	46
44	42	47	43	46	46	34	43	45	39	47	45	46	45	46	45
44	38	46	39	42	43	33	48	41	40	43	44	40	41	42	41
43	41	44	42	45	45	34	40	46	39	46	46	43	47	45	46
43	42	45	44	47	46	35	41	48	39	47	48	43	46	46	48
43	41	46	42	45	47	35	42	46	40	46	46	43	46	47	46
39	44	42	45	41	42	34	39	40	34	42	40	40	40	41	40
44	43	46	45	48	45	34	42	47	38	48	47	44	45	45	47
42	42	44	44	46	46	33	43	45	36	46	45	45	43	43	45
40	39	43	40	42	47	36	42	43	37	43	43	43	43	44	43
43	39	42	40	43	43	32	41	44	38	44	46	41	45	43	44
42	41	46	42	46	44	34	39	46	38	46	46	42	45	44	47
X	38	42	39	45	41	30	42	41	36	43	43	44	42	41	44
1.893	X	40	45	42	40	30	37	40	32	45	40	39	40	39	41
1.124	1.452	X	41	44	44	33	44	43	39	45	43	42	43	44	43
1.669	0.363	1.241	X	43	41	31	38	42	33	43	42	40	40	40	42
0.628	1.124	0.817	0.924	X	43	33	40	45	36	47	45	45	43	43	48
1.33	1.452	0.73	1.241	1.012	X	34	43	44	38	44	44	44	44	45	44
3.389	3.289	2.551	3	2.649	2.291	X	32	34	30	33	33	31	33	33	34
1.213	2.214	0.817	1.983	1.719	1.012	2.918	X	39	40	41	42	41	39	40	39
1.419	1.541	1.012	1.124	0.714	0.817	2.388	1.937	X	37	45	46	41	44	45	46
2.361	3.301	1.669	3.03	2.453	1.893	3.289	1.541	2.214	X	37	38	34	38	38	37
1.1	0.628	0.714	1.012	0.433	0.904	2.745	1.594	0.8	2.304	X	45	43	45	44	46
0.924	1.452	0.924	1.035	0.628	0.73	2.551	1.213	0.442	1.893	0.714	X	41	44	44	46
0.641	1.578	1.035	1.361	0.54	0.641	3	1.33	1.33	2.768	1.012	1.241	X	42	41	44
1.124	1.452	0.924	1.452	1.012	0.73	2.551	1.848	0.817	1.893	0.714	0.73	1.035	X	44	44
1.33	1.669	0.73	1.452	1.012	0.54	2.551	1.63	0.628	1.893	0.904	0.73	1.241	0.73	X	44
0.817	1.33	1.012	1.124	0.172	0.817	2.388	1.937	0.529	2.214	0.614	0.442	0.73	0.817	0.817	X

79C

3.1.2 – Southern analysis of *hyl* PCR Negative Mutants

Several of the strains included in the study were PCR negative for hyaluronate lyase, but PCR positive for *hexA* / *recP*. Given the PCR controls used, and the successful amplification of other genes from the DNA of these strains, it was decided to screen them for *hyl* by Southern analysis, using the digoxygenin labeling and detection protocols (see chapter 2).

hyl allele 2 was amplified from strain 916, and as it was the most commonly represented allele. This was used to probe *Kpn*I and *Msp*I chromosomal digests of the PCR negative strains, along with R6 and strain 471 (also allele 2 positive).
The results of the *Kpn*I blot are shown in figure 3.1.1 (the *Msp*I blot also proved positive for the isolates used).

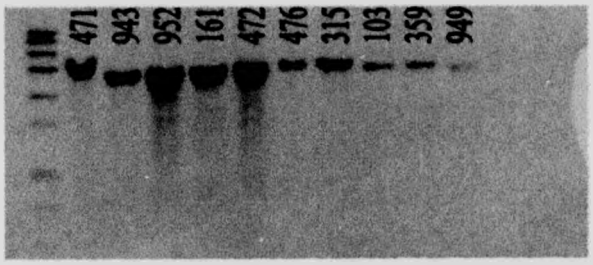


Figure 3.1.1 – Southern blot of PCR negative strains probed with the *hyl* PCR product. Markers in lefthand lane = Dig labelled SPP1 cut with *Eco*RI. Strain ID numbers are used. Strains 359 and 949 were included as positive controls.

As can be seen, all strains were positive for *hyl* by Southern analysis. Five of these strains were included in the phenotypic analyses described in chapter 3.2, and proved to be Hyl^+.

3.1.3 – Comparison and interpretation of HRRA results

3.1.3.1 - *recP*

Figure 3.1.2 shows the number of strains that possess each of the 17 *recP* alleles. As can be seen, the most common allele is allele1, followed by allele 6, then allele 5. These groups of alleles represent the three major allele types seen with this locus.

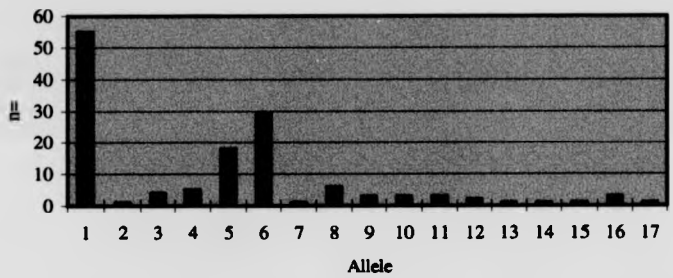


Figure 3.1.2 - *recP* allele occurrence.

Within this group, the maximum estimated percentage divergence was 2.607% between alleles 2 and 15. Both of these alleles occur once, in strains 18B (ID922) and 142163 (ID324) respectively. The maximum estimated percentage divergence from the three major allele representatives is given in table 3.1.5.

Common Allele	Diverged Allele	Max. % Divergence	Strain
1	14	1.893	619 (ID53)
5	14	2.232	619 (ID53)
6	13	2.007	54 (ID967)

Table 3.1.5 - Maximum estimated percentage divergence from the three most common *recP* alleles found.

Of the 17 pneumococcal alleles found the distribution of allele types within each isolate group (carried, invasive and serotype 3 isolates) varies as follows:

The carried strains were found to possess 13 different allele types, of which alleles 2, 7, and 13 are unique to that group. In this study efforts have been made to exclude the possibility of studying a number of clonal isolates (such as varying serotype, site and date of isolation and geographical origin). Where the number of alleles unique to a specific group of isolates is large, it is arguable that there is some selective pressure to maintain those alleles within that group. The actual number of carried strains which contain unique alleles represents a small proportion of the total number of strains studied (5/140 or 3.57%), as such this cluster of unique alleles is not thought to suggest any such selection. χ^2 tests were used to analyse the distribution of unique and shared alleles for the genes examined across the groups of carried, invasive and serotype 3 strains. In all cases the number of strains bearing unique alleles was not found to vary significantly, with $\chi^2 < \chi^2_{crit}$ ($\chi^2_{crit} = 5.99$).

The invasive isolates contain 13 different alleles, of which alleles 14, 15, 16 and 17 are unique to the group. These represent 4.28% of the total number of strains (6/140), and again, are not thought to represent allele selection.

The serotype 3 isolates all possess *recP* alleles seen in either the carried or invasive isolates.

3.1.3.2 – *hexA*

Figure 3.1.3 shows the distribution of *hexA* allele frequencies.

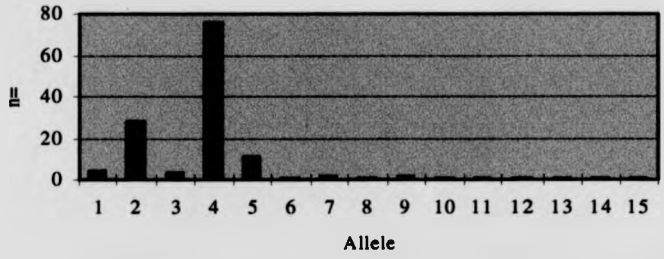


Figure 3.1.3 - *hexA* allele occurrence.

The *hexA* alleles found in this study number 15 in all. Of these, 107/138 (77.5%) strains possess one of two major representative alleles (allele 2 or 4). The estimated percent divergence between these two alleles is 0.389%. The maximum estimated percent divergence from these alleles is shown in Table 3.1.6.

Common Allele	Diverged Allele	Max. % Divergence	Strain
2	12	2.592	142154 (ID326)
4	12	1.994	142154 (ID326)

Table 3.1.6 - Maximum estimated percent divergence from the two predominant *hexA* alleles.

Thus allele 12, from strain 142154 (ID326) is the most divergent from the majority of other strains. It is a serotype 12 Kenyan isolate, taken from a patient suffering from invasive pneumococcal septicaemia. The other loci in this strain (*recP* and *hyl*) contain non-unique alleles.

The carried isolates contain 11 different alleles, of which 6, 7, 8 and 10 are unique to that group. These represent 3.62% (5/138) of the total number of

The carried isolates contain 11 different alleles, of which 6, 7, 8 and 10 are unique to that group. These represent 3.62% (5/138) of the total number of strains from which *hexA* was amplified. Using the argument outlined above, it is not thought that this represents underlying biological selection.

Of the alleles found in the group of invasive strains, alleles 12, 13, 14 and 15 were unique to the group. The relative percentage of strains in this 'unique' set of alleles is 2.89% (4/138).

The serotype 3 collection all possessed either allele 2 or 4, the predominant alleles over all the strains included. Examination of the distribution of housekeeping genes within the type 3 group, and the relatively restricted number of *hexA* alleles indicates that this group is not highly diverged, and probably consists of a discrete number of clonal elements.

This distribution of unique allele representation across the 3 groups studied was not significant by the χ^2 test ($\chi^2 = 3.32$, $\chi^2_{crit} = 5.99$).

3.1.3.3 – Hyaluronate lyase - *hyl*

133 strains were PCR positive for *hyl*, and found to posses 28 different alleles as determined by HRRA. Figure 3.1.4 shows the relative frequencies of these alleles.

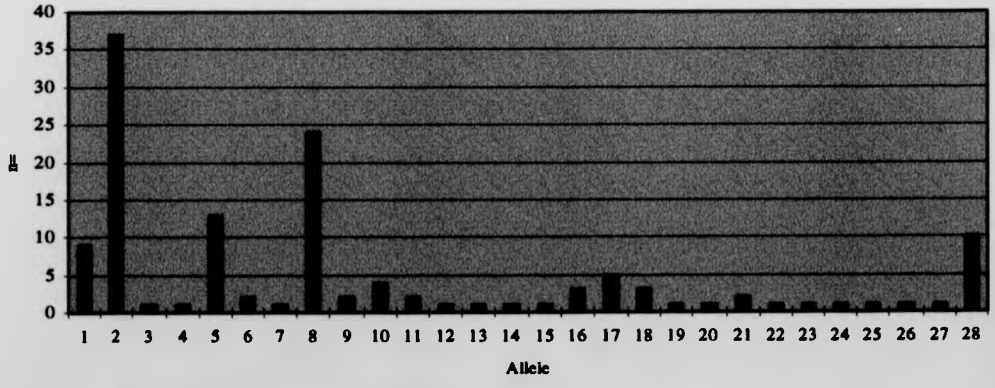


Figure 3.1.4 - *hyl* allele occurrence.

Four of the predominant *hyl* alleles (1,2, 5 and 8) are found distributed throughout the groups of isolates from nasopharyngeal carriage or invasive disease, and alleles 1, 5 and 8 show no association with one group or another as determined by the χ^2 test.

Allele 2 is the most frequently represented. This representation may contain a bias from the inclusion of 10 serotype 3 isolates from the UK (some of which were isolated from the same town) that contain common *hexA* and *recP* alleles along with *hyl* allele 2. This is supported by the χ^2 test, which gives a value of $\chi^2 = 11.162$ ($\chi^2_{crit} = 5.99$).

Muller-Graf *et al.* described a nationally distributed serotype 3 clone on the basis of HRRA data obtained from HRRA data from 9 housekeeping genes (Muller-Graf *et al.*, 1999). It is possible that these isolates may be part of this clone, although the power of resolution of the HRRA data obtained here (ie; the ability to clearly identify distinct clonal relationships) is not high enough to clearly identify clonal groups (using 3 loci, as opposed to 9). The association of allele 2 with what may be a successful national clone, and its global distribution across a number of strains of *S. pneumoniae* of different serotypes and genotypes (*hyl* allele 2 is found in the context of 16 discrete combinations of *hexA* and *recP* alleles in this study) suggests that allele 2 is itself a successful allele.

Strains with *hyl* allele 28 (n = 10) are the equine isolates derived from a recent study carried out in our laboratory (Whatmore *et al.*, 1999). These strains also possess common *recP* and *hexA* alleles, supporting their clonal relationship.

During the PCR amplification of the *hyl* products, one strain (619, ID53) gave rise to an amplicon approximately 400bp smaller than those from all of the other PCR positive strains. It was assumed that this size difference was as a result of a naturally occurring deletion event. This strain was kept in the study, and was further characterised as described in section 3.3.2.

The maximum estimated percentage divergence between the predominant alleles is shown in table 3.1.7 (allele 19, assigned to the deletion mutant was excluded from this analysis).

Common Allele	Diverged Allele	Max. % Divergence	Strain
1	22	2.214	86012 (ID307)
2	22	1.759	86012 (ID307)
5	22	1.848	86012 (ID307)
8	22	2.073	86012 (ID307)
28	22	2.214	86012 (ID307)

Table 3.1.7 - Maximum estimated percent divergence between the five predominant *hyl* alleles.

It is clear from table 3.1.7 that of the most common alleles the most highly diverged is allele 22, followed by 7, 14 and 20.

3.1.3.4 – Discussion

The results of the HRRA approach presented here indicate that variation at the *hyl* locus is not notably higher than that observed for loci encoding two housekeeping genes, *recP* and *hexA*. This was shown using the χ^2 test comparing the *hyl* maximum estimated percentage divergence ($\%_{max}$) and the variation of the maximum and minimum estimated percentage divergence ($\%_{max}$ -$\%_{min}$) with the values obtained for *hexA* and *recP* alone, or and average percentage divergence value obtained for the two housekeeping genes taken together. In all cases $\chi^2 << \chi^2_{crit}$.

Despite the fact that hyaluronate lyase is likely to be located at the cell surface (see section 1.7.10, chapter 1), the *hyl* locus does not appear to have been under strong selective pressure to diverge, as might be expected. However, it is interesting that there are a higher number of distinct *hyl* alleles than for the housekeeping genes. This difference in allele numbers was shown to be statistically significant by the χ^2 test (χ^2 = 10.9, χ^2_{crit} = 5.99). This feature will be returned to in section 3.1.6.1.

The estimated percentage divergence at the *hyl* locus is lower than that of other extracellular virulence determinants, including *nanA* (36% - S.King, unpublished data), *lytA* (3.2% - Whatmore and Dowson, 1999) and *iga* (12% - Poulsen *et al.*, 1998).

The HRRA/Nei technique relies upon the detection of polymorphisms within a defined restriction endonuclease 'window' to obtain an estimate of the overall

variation at a given locus. As a result of which there will be background variation at the nucleotide level which remains undetected (i.e; polymorphisms that do not result in restriction site alterations). One element of the Nei and Li treatment is to account for this 'silent' variation.

Representative *hyl* alleles were sequenced to complement the HRRA data, and to determine the nature of nucleotide variation at this locus.

3.1.4 - Sequence analysis of selected *hyl* alleles

The HRRA technique is designed to obtain an estimate of the percentage divergence values expected at the nucleotide level. Since it relies upon the restriction recognition footprint of the enzymes used to generate the raw data, it can be regarded as a form of low-resolution sequence analysis.

In order to verify the validity of the HRRA technique, and to examine the nature of the variation at the nucleotide level, nine of the alleles were selected for direct sequence analysis over the hyaluronate lyase ORF. These were the predominant 4alleles (1, 2, 5, 8 and 28) and those which were most highly diverged from them (alleles 7, 14, 20 and 22). Alleles were sequenced from the following strains; allele 1; strain ID 885, allele 2; strain ID 922, allele 5; strain ID 903, allele 7; strain ID 873, allele 8; strain ID 877, allele 14; strain ID 863, allele 20; strain ID 303, allele 22; strain ID 307, allele 28; strain ID 498.

The region selected for sequence analysis corresponds to nucleotides 113-2978 of the published *hyl* sequence (accession no. L20670) (Berry *et al.*, 1994), corresponding to amino acids 9-stop codon, plus 12 codons downstream of this.

Sequencing was carried out using the Beckman CEQ-2000 system and ABI Prism system as described in chapter 2.

3.1.4.1 - Observed alterations to the published sequence

Sequences were aligned using the Clustal alignment algorithm contained within the Megalign program of the DNAStar suite. Alterations to the published

nucleotide sequence were confirmed from data obtained by sequencing the complementary DNA strand.

The occurrence of restriction sites used in the HRRA study within the sequences of the nine alleles selected confirmed the polymorphisms leading to the observed restriction types and hence alleles derived in section 3.1.1.4.
The observed nucleotide alterations are summarised in figure 3.1.5, below.

```
                                1111111 1111111111 1112222222 22222222222 222222
      1112333334 4555666677 7790001111 2223346779 9990001111 12224455566 677799
      4568678894 7269014734 5932681578 2482709561 2692220227 88898938915 727901
      8480511395 6411775820 8172239545 5575385198 6660199148 42354550471 016774

al1   ....CT..GT .G....TA.. C.....TA.. ....C.T... ..CGTAGCT. ATC...T... ..T. ..
al2   ..G.....GT G...G..AT. CT........ .G........ .....A...C ATC..GT.A. .C.. ..
al5   C.G..T..GT ....GC.AT. CT....TA.. ..T..T.... .....A.C.. ATCACGT... .C.. ..
al7   C..A..CAGT ........A.. .......... .......... .....A.... ATC...T... .... ..
al8   C....T..GT .G..G..A.. C.C...TA.. ....G..... .....A...C .TC...TT.. .... ..
al14  C.....CAGT ....G..ATT CT....TA.. ..T..... .....A...C .TC...TT.. .... ..
al20  .....T..GT ...GG..A.. C...ACT... .G.CG..... ........C.. .TC.C.T... .... .C
al22  .A..CT..GT ...GC..A.. C..T...... .G..C..TGT ..C..AGC.C .TC.CGT... .C..V..
al28  C.G.....GT ..A....A.. ......T.AA A.T.G..... CGC...GC.. .TC.C.TT.A .C.T AC
L20670 AGAGACTCAC AAGAAACCCC TCAAGTCGTG CCCTACCCAC TATAGGATCT GGGGTACCGT TTCC GT

P.type CGAGATTCGT AAGAGACACC CCAAGTTGTG CCCTGCCCAC TATAGAACCT GTCGTATCGT TTCC GT
```

Figure 3.1.5 – Summary of nucleotide alterations found in the 9 alleles studied. The published sequence (accession number L20670) is shown in blue, and the prototype sequence derived from this study in black (P. type). The numbers running vertically indicate positions of polymorphic sites with respect to L20670. Bases shown in red are variant with respect to L20670 in all the cases studied. Bases indicated in purple are variant in over 50% of cases. ∇ represents a 24 bp duplication found in allele 22. This is discussed in the text below.

The nucleotide sequences of the alleles studied were translated into their respective proteins, and an alignment of these protein sequences was also made. Non-synonymous mutations leading to amino acid alterations are shown in figure 3.1.6, below.

```
                  111 111 222 233 334 445 566 667 888 889          9
            226 999 124 779 123 816 781 245 612 443 034 590          0
            024 248 995 366 734 312 952 804 006 471 253 463          4
al1         ... AL. S.G ..E .T. ... ... S.. ... VNV ... ..
al2         ... ... SS. A.E .TM ... .E. ... ... .NV G*. ..
al5         H.. .L. S.. AAE .TM ... ... .Y. ... .NV G.. ..
al7         H.I ..Y S.. ..E ... ... ... ... ... .NV ... ..   Duplicated
al8         H.. .L. S.G A.E .T. R.. ... S.. ... .NV ... ..   Region
al14        H.. ..Y S.. A.E LTM ... ... ... ... .NV ... ..
al20        ... .L. S.. A.E .T. ... .ES S.. ... ..V ... ..
al22        .I. AL. S.. A.E .T. .V. .E. S.V RS. .NV G.. ..RNLKRKKK
al28        H.. ... S.. ..E ... ..I K.. S.. ..E ..V ..T .S
L20670      NVV DPS PNE TDD SIT SEL QDL NHA KPK SDG DWS LP- --- --- -

Prototype   HVV DLS SNE ADE STT SEL QDL SHA KPK SNV DWS LP
```

Figure 3.1.6 – Summary of amino acid alterations found in the 9 alleles studied. The translated published sequence (accession number L20670) is shown in blue and the prototype derived from this study in black. The numbers running vertically indicate positions of polymorphism with respect to L20670. Residues indicated in red are variable with respect to L20670 in all the cases studied. Residues indicated in purple are variable in over 50% of cases. The 8 amino acid duplication found in the translation product of allele 22 is shown. This is discussed in section 3.1.6. * indicates amber stop codon.

Table 3.1.8 summarises the observed nucleotide and predicted amino acid alterations, and includes the ratios of synonymous to non-synonymous changes. It also includes the number of nucleotide alterations that result in changes to the restriction type of each given allele.

Allele	No. nt changes		No. aa substitutions		S/NS ratio	% nt changes that lie within a restriction site with respect to P. type
	w.r.t L20670	w.r.t P.type	w.r.t L20670	w.r.t P.type		
1	24	14	10	5	1:0.41	14.2
2	20	15	12	8	1:0.6	6.6
5	25	13	12	5	1:0.48	0
7	12	10	7	6	1:0.58	16.6
8	18	6	11	2	1:0.6	0
14	20	12	10	5	1:0.5	16.6
20	19	9	9	4	1:0.47	11.1
22*	26	16	16	9	1:0.61	12.5
28	27	21	10	9	1:0.37	9.5

Table 3.1.8 – Statistics relating to the alterations of the nine alleles studied. Alterations were scored with respect to the prototype sequence given in figure 3.1.6. nt = nucleotide, aa = amino acid, s/ns = synonymous / non-synonymous, w.r.t = with respect to..., P. type = prototype sequence. *Note that allele also contains a 24bp imperfect duplication as well as the other observed alterations.

The observed nucleotide divergence is shown in table 3.1.9, below. The values estimated by the Nei and Li algorithm are also included. The table also includes amino acid identity/similarity data for the transcribed protein.

Allele	1	2	5	7	8	14	20	22	28
1	X	**0.907**	**0.803**	**0.698**	**0.558**	**0.837**	**0.733**	**0.768**	**1.012**
		0.714	0.257	1.241	0.433	1.33	1.3	2.214	0.529
2	98.935	X	**0.523**	**0.558**	**0.628**	**0.558**	**0.733**	**0.698**	**1.082**
	99.255		0.433	0.836	0.257	1.124	1.1	1.759	0.714
5	98.937	99.361	X	**0.663**	**0.593**	**0.523**	**0.698**	**0.803**	**0.907**
	99.256	99.574		1.124	0.168	1.213	1.594	1.848	0.257
7	99.044	99.042	99.044	X	**0.488**	**0.419**	**0.663**	**0.837**	**0.872**
	99.362	99.468	99.256		0.924	0.462	1.669	2.674	1.452
8	99.469	99.042	99.256	99.150	X	**0.349**	**0.523**	**0.698**	**0.872**
	99.575	99.468	99.469	99.362		1.012	1.387	2.073	0.433
14	98.937	99.361	99.362	99.469	99.469	X	**0.733**	**0.907**	**0.942**
	99.150	99.681	99.469	99.575	99.362		2.214	3.301	1.33
20	99.256	99.255	99.044	98.937	99.362	99.044	X	**0.593**	**0.837**
	99.575	99.468	99.469	99.362	99.787	99.362		1.541	1.937
22	98.937	98.935	98.725	98.406	98.831	98.512	99.044	X	**1.012**
	99.469	99.148	99.150	99.044	99.469	99.044	99.469		0.817
28	98.725	98.509	98.512	99.044	98.831	98.725	98.831	98.087	X
	99.150	99.042	98.831	99.362	99.150	98.937	99.150	98.831	

Amino acid

Table 3.1.9 – Percentage divergence of selected *hyl* alleles at the nucleotide level. The upper right part of the matrix shows percentage nucleotide divergence in bold, as calculated by the Needleman and Wunsch algorithm from direct sequence data. The values in grey are those derived from the algorithm of Nei and Li using the data generated by HRRA. The lower left part of the matrix shows the amino acid percentage identity (green) and similarity (red) as derived from the translated nucleotide data.

As can be seen from the table above, the overall divergence at the nucleotide, and hence amino acid level, is low. The estimated percentage divergence as calculated by the Nei and Li algorithm is in some cases, broadly accurate. However, in many cases the estimates prove to be either higher or lower than that observed from direct sequence analysis. Analysis of the mean divergence data obtained by each method (DNA sequencing and HRRA) were compared using a paired t-test. This indicated that the data sets were significantly different ($t = 4.33$, $t_{crit} = 2.03$, $P<0.001$). Possible reasons for this difference are discussed in section 3.1.4.3, below.

Phylogenetic analysis of the sequence data was carried out in order to generate dendrograms from both the nucleotide and amino acid sequences. These are presented in figure 3.1.7, below.

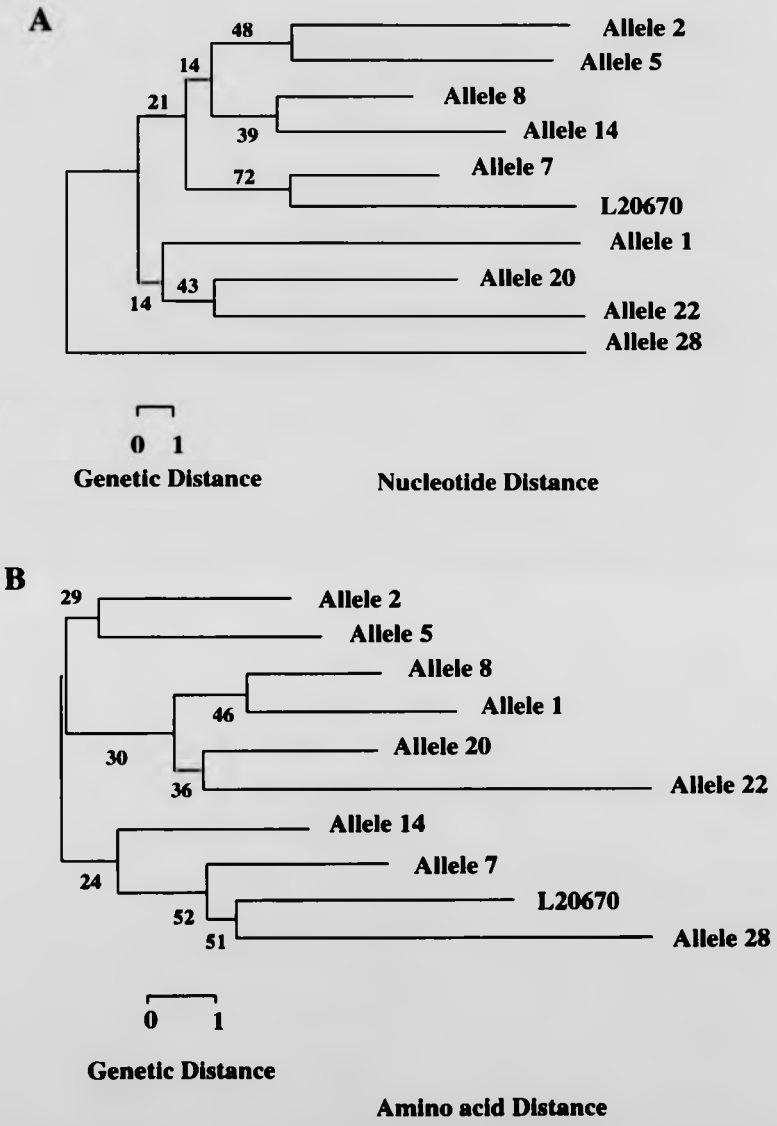


Figure 3.1.7 – Phylogenetic relationships between the *hyl* alleles sequenced and predicted translation products. Analysis was made using the Neighbour-Joining method. Bootstrap values were calculated on the basis of 500 replicates. Analysis was carried out using nucleotide (A) and amino acid (B) sequence alignments.

As can be seen from dendrogram A shown above, the relationships between the sequenced alleles reflect the nucleotide divergence shown in the data matrix given in table 3.1.9. The difference between the trees indicates the differing occurrence of non-synonymous changes between the alleles sequenced. Since the overall number of alterations is low, and the nucleotide/amino acid stretch is large, the reliability is such analysis is open to question, as is reflected by the low bootstrap values at all of the assigned breakpoints. As such, limited emphasis can be placed on the associations drawn from such analysis.

It is noteworthy that there is no obvious clustering of either common or diverged alleles. This suggests that there has been no partition of particular alleles on the basis of the phenotype conferred upon the respective isolates from which they originate.

3.1.4.2 – Distribution of observed alterations

An interesting feature of the sequence data obtained relates to the distribution of alterations over the length of the gene studied. Analysis of the distribution of nucleotide and amino acid alterations indicates that there is more amino acid variation in the α-domain of the protein (which includes 55% of the amino acids). Table 3.1.10 shows the percentage of nucleotide alterations and amino acid substitutions in the region encoding the α- domain of the protein, and the percentage of amino acid alterations in these domains.

Allele	% nt changes in α-domain	% aa changes in α-domain
1	50	80
2	60	75
5	61	80
7	80	100
8	50	100
14	75	100
20	66	75
22	43	55
28	43	55

Table 3.1.10 – Distribution of nucleotide and amino acid alterations in the nine alleles examined. Alterations were scored with respect to the prototype sequence given in figure 3.1.6. nt = nucleotide, aa = amino acid.

Although the α-domain comprises slightly more (55%) of the molecule than the β-domain in terms of amino acids, in all cases there appears to be a bias in the distribution of amino acid alterations toward the α-domain. This is interesting as it contains the catalytic residues. This feature will be returned to in section 3.1.5.2

3.1.4.3 – Discussion

Table 3.1.9 shows that in some cases the percentage nucleotide divergence predicted by the algorithm of Nei and Li is either higher or lower than that observed from direct sequence analysis. The extent of disagreement between these values reaches significance as determined using a paired T-test ($t = 4.33$, $t_{crit} = 2.03$, $P<0.001$).

The Nei algorithm was developed with various starting assumptions, including the stochastic principle of molecular evolution; that nucleotide variation is distributed randomly over the length of DNA studied. As a result, the observed restriction site changes might also be expected to vary randomly. When variation occurs in a fashion, which differs from such an equal distribution, such analysis begins to break down.

While the overall extent of variation is low, such that phenomena such as homologous recombination between divergent alleles is not immediately apparent, other sources of error must be sought to explain the disparity between the predicted (by HRRA) and observed nucleotide divergence.

It is suggested here that these differences can be understood by looking at the distribution of nucleotide changes over the alleles sequenced in the following manner:

Table 3.1.8 shows the relative percentage of polymorphisms that fall within the overall restriction footprint of the restriction endonucleases used in the HRRA procedure. In many cases where the Nei and Li algorithm has overestimated the percentage divergence, a relatively high proportion of the total number of polymorphisms for a given allele fall within the restriction endonuclease footprint.

This is the case with alleles 7 and 14 for example, in which 16.6% of all polymorphisms alter their restriction type with respect to the prototype sequence. The Nei and Li algorithm has overestimated the percentage nucleotide divergence in 8/8 comparisons made with these alleles. A similar situation is seen with allele 22, where 12.5% of polymorphisms fall within a restriction site, and divergence is overestimated in 8/8 comparisons. This interpretation does not hold for allele 1 however, where 14.2% polymorphisms affect the restriction type, but the Nei and Li algorithm has overestimated divergence in only 4/8 comparisons.

The converse situation is less readily explained. Alleles 5 and 8 contain no polymorphisms that alter their restriction type, but the Nei and Li algorithm has underestimated the percentage divergence in only 4/8 comparisons. This is explained as follows; these alleles vary in such as way as to result in no changes to their restriction types, so no information is gained pertaining to their variation with respect to the prototype sequence. This renders the assessment of variation within these alleles as a function of the stochastic occurrence of restriction type 'active' polymorphisms within the alleles to which they are being compared.

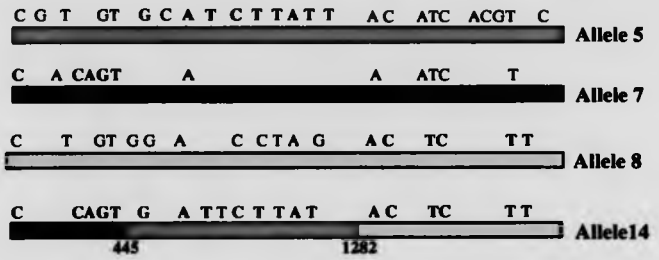
This interpretation of the data suggests that while over-representation of polymorphisms in the restriction footprint can lead to a positive bias in the output of the Nei and Li algorithm, under-representation does not.

3.1.5 – Searching for recombination between *hyl* alleles

Visual analysis of the distribution of nucleotide alterations between the alleles shown in figure 3.1.5 indicates that blocks of alterations are apparent across the sequences studied, however Sawyers Runs tests indicated that these putative recombination events are not statistically significant (Sawyer, 1989). It

is suggested that this is due to the relatively low number of polymorphisms over a large surface of comparison.

Maximum Chi-squared tests were also performed (Maynard Smith, 1992). This technique has the advantage that individual putative points of recombination can be submitted for statistical analysis. In this case recombination was shown to occur between the alleles identified, and probability scores reached statistical significance ($P<0.0100$). Two examples of recombination events shown to be statistically significant are shown in figure 3.1.8, below.



Or the more complicated example of allele 22, which bears polymorphic features similar to a number of other alleles;

Figure 3.1.8 – Mosaic structures within *hyl* alleles as determined by the maximum χ^2 test. In all cases putative recombination points are statistically significant, with $P \leq 0.0100$.

Thus, there is statistical evidence that the pneumococcal *hyl* locus is evolving by transfer of blocks of DNA of low overall divergence, leading to the formation of detectable mosaic gene structures.

Recombination at the *hyl* locus could be occurring in two ways. The accumulation of point mutations followed by intra-species recombination may lead to the structures observed. Alternatively, divergent DNA could be exchanged at the inter-species level.

The polymorphic regions of the pneumococcal *hyl* alleles identified by sequence analysis were compared with homologous regions from the hyaluronate lyase genes of other bacteria available in the public domain. This investigation did not indicate that there were any obvious

potential donors for divergent DNA. This is unsurprising in the light of the low overall divergence observed within the pneumococcal *hyl* locus, and the relatively high divergence of other streptococcal hyaluronate lyase encoding genes available in the public domain (≈ 42% nucleotide divergence between the pneumococcal *hyl* locus and its homologues in both *S. agalactiae* and *S. pyogenes*). If genetic exchange with such donors were taking place, relatively highly divergent blocks of DNA would be apparent in the pneumococcal alleles sequenced. As such it was decided to conduct searches for potential homologues, and hence donors in more closely related streptococci.

Although *S. constellatus* and *S. intermedius* are hyaluronate lyase positive, Southern analysis of genomic DNA from these organisms proved either negative or only weakly positive when probed with *hyl* DNA from *S. pneumoniae* (S. Widdison - unpublished data).

Previous studies have shown that pneumococcal genes encoding PBP's (Dowson *et al.*, 1993), LytA (Poulsen *et al.*, 1998) and NanA (S. King – unpublished data) have been involved in recombination with loci from the oral streptococci *Streptococcus oralis* and *S. mitis*. Furthermore, a recent study by Whatmore *et al.* has revealed that certain groups of 'atypical' isolates of *S. mitis* harbour the pneumococcal virulence determinant genes *ply* and *lytA* (Whatmore *et al.*, 2000). Also, members of these groups of organisms were positive in Southern analyses of Widdison (Unpublished data).

As such it was thought that these closely related groups of isolates would be good candidates for donors of divergent DNA which might shed light on the polymorphisms seen at the pneumococcal *hyl* locus.

3.1.5.1 – Screening other streptococci for *hyl* homologues

The approach taken here was based on specific and degenerate PCR methodology. The primer pairs used in this screen were the hyl and 2°hyl probe (hyp 2°) primers described in table 2.7, chapter 2. The latter primer pair contain degenerate sites, designed on the basis of homology to the group A and group B streptococcal *hyl* sequences. These primers successfully amplified products from representatives of the pneumococcal strains that were previously identified as PCR-negative in the HRRA study (section 3.1.2). The strains used in this experiment are listed in chapter 2, table 2.6.

Figure 3.1.9 (below) shows the relative annealing positions at the *hyl* locus, with the numbering of the available database sequence of the gene used throughout (accession number L20670). Also included in this figure is a photomontage

some of the PCR positive and negative results obtained. Table 3.1.11 contains a full description of the results obtained with the strains used.

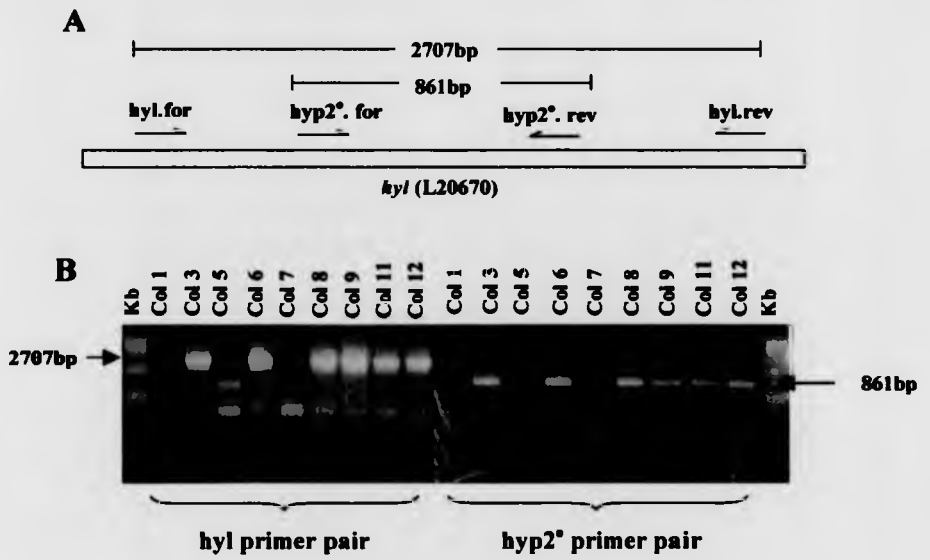


Figure 3.1.9 – PCR screen for *hyl* amongst typical and atypical pneumococci, and other oral streptococci. A; PCR rationale, B; Products separated through 1% agarose.

As can be seen, those strains that have given rise to amplification of a product of the expected size with the specific hyl primer pair have also done so with the degenerate hyp2° primer pair. Some templates gave rise to an amplification product that was smaller than expected (Col 5/hyl for example).

The results obtained from all strains tested are shown in table 3.1.11.

Strain	Strain ID	Species	Group*	hyl primer pair	hyp2° primer pair	*hyl* positive?
Col 1	667	*S. pn* (atypical)		-	-	No
Col 3	668	*S. pneumoniae*	A	+	+	Yes
Col 5	669	*S. pn* (atypical)		800	-	No
Col 6	670	*S. pneumoniae*	A	+	+	Yes
Col 7	671	*S. pneumoniae*	A	-	-	No
Col 8	672	*S. pneumoniae*	A	+	+	Yes
Col 9	673	*S. pneumoniae*	A	+	+	Yes
Col 11	674	*S. pneumoniae*	A	+	+	Yes
Col 12	675	*S. pneumoniae*	A	+	+	Yes
Col 13	676	na		+	+	Yes
Col 14	677	*S. pneumoniae*	A	+	+	Yes
Col 15	678	*S. mitis*	B	1000	-	No
Col 16	679	*S. mitis.*	B	-	+	No
Col 17	680	*S. mitis* (atypical)		-	+	No
Col 18	681	*S. mitis*	B	-	-	No
Col 19	682	*S. mitis*	B	600	-	No
Col 20	683	*S. mitis*	B	-	-	No
Col 21	684	*S. oralis*	C	600	-	No
Col 22	685	*S. mitis*	B	-	-	No
Col 24	686	*S. pn* (atypical)		-	-	No
Col 25	687	*S. oralis*	C	700	-	No
Col 26	688	*S. pn* (atypical)		-	-	No
Col 27	689	*S. pn* (atypical)		-	-	No
Col 28	690	*S. mitis*	B	-	-	No

Table 3.1.11 – PCR screen for *hyl* in related streptococci. Where bands were observed, but not of the expected molecular weight, approximate sizes are given. Species designations of typical and atypical are based on the findings of Whatmore *et al.* (Whatmore *et al.*, 2000). na - not available.

In this study, those strains that gave rise to a positive band of the expected size in both sets of PCR reactions were deemed *hyl* positive. Those that gave rise to only one band, or bands of spurious sizes were said to be *hyl* negative.

As can be seen, only those isolates determined to be typical pneumococci by Whatmore *et al.* proved to be *hyl* positive in this screen. One isolate (Col 7) belongs to an outgroup of the typical pneumococci, along with Col 11 and Col 12. Interestingly, Col 7 was determined to be *hyl* negative, whereas Col 11 and Col 12 were *hyl* positive.

Many of the strains examined in this study have previously been shown to harbour *ply* and *lytA*. Data presented here suggest that these strains do not contain *hyl* or loci bearing close similarity to *hyl*. This indicated a rather striking restriction of this locus to those strains identified as 'typical' pneumococci. This area is interest is under active investigation by another member of the group, following up these experiments with a Southern based screen for *hyl* homologues.

3.1.5.2 – Discussion

The preceding sections show that the overall nucleotide divergence at the *hyl* locus is low, and that there is a bias for polymorphisms resulting in amino acid alterations to the translated protein to the α-domain. This may reflect either greater freedom of steric variation in the α-domain compared to the β- domain, or may have implications in terms of the antigenicity of this part of the protein.

If the α-domain proved to harbour more immunodominant epitopes, then selective pressure from host immune surveillance might result in more alterations to this domain. If this were the case, we might expect to find the altered residues more frequently found in surface exposed loops, linking structural elements of this domain.

The crystal structure of pneumococcal hyaluronate lyase has recently been reported (Li *et al.*, 2000). Detailed information regarding the precise location of the individual α-helical components of the α-domain would enable us to precisely map the observed alterations to the protein, and further hypothesise on this issue. Representations have been made to the authors regarding this, however no additional information has been forthcoming.

Results described in section 3.1.5 indicate that there is statistical evidence for recombination between pneumococcal *hyl* alleles. The inability to map correlate the polymorphisms observed with the homologous regions of other hyaluronate lyase sequences available in the public domain, or to locate homologous loci in related streptococci suggests that the variation observed is restricted to the intra-species level. The occurrence of recombination between alleles of low overall nucleotide divergence, and the observation that polymorphisms are

may have some bearing on the relatively high number of discrete *hyl* alleles found by HRRA, as compared to *hexA* and *recP*.

Another issue concerns the canonical pneumococcal sequence of *hyl* deposited by Berry *et al.* under the accession number L20670 (Berry *et al.*, 1994). It is clear from the polymorphic sites represented in figure 3.1.5 that some modifications to the consensus sequence are required. These are given as follows, with amino acid alterations given in brackets. Numbering is shown with respect to L20670. $A^{399}\rightarrow G$, $C^{445}\rightarrow T$ ($P^{119}\rightarrow S$), $C^{677}\rightarrow A$ ($D^{196}\rightarrow E$), $G^{2281}G^{2282}\rightarrow TC$ ($G^{731}\rightarrow V$). Alterations that occur in more than 50% of observed cases (5/8 or greater) are $A^{147}\rightarrow C$ ($N^{20}\rightarrow H$), $C^{370}\rightarrow T$ ($P^{94}\rightarrow L$), $A^{606}\rightarrow G$ ($T^{173}\rightarrow A$), $T^{757}\rightarrow C$ ($I^{223}\rightarrow T$), $C^{1118}\rightarrow T$, $A^{1372}\rightarrow G$ ($N^{428}\rightarrow S$), $G^{2028}\rightarrow A$ ($D^{646}\rightarrow N$) and $T^{2120}\rightarrow C$.

3.1.6 – Polynucleotide duplication in allele 22

Direct sequence data obtained in section 3.1.4 revealed that allele 22 (sequenced from a serotype 19 carried isolate from Kenya – ID 307) apparently contains a 24bp duplication near its 3' end. Analysis of the region containing this duplication indicates that two independent duplication events seem to have occurred. This is illustrated in figure 3.1.10.

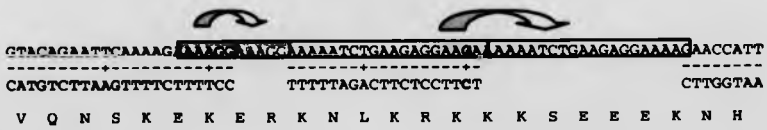


Figure 3.1.10 – Duplicated region found in allele 22. The duplicated region is shown in red typeface, with the proposed duplications indicated.

As can be seen, the inserted sequence consists of two discrete elements, one 4 bp in length and one of 20 bp. These lie in the region encompassing bases 2797-2821 of the L20670 sequence, with a single A → G polymorphism in the 20 bp reiterated sequence. The structure observed seems to have arisen from two

reiterated sequence. The structure observed seems to have arisen from two separate duplication events, either one alone would result in truncation of the translated protein. Taken together, these insertions restore the wild type protein sequence after the 20 bp insertion.

3.1.6.1 – Discussion

The nature of this region is interesting, especially in the light that together, these sequences restore the reading frame of this mutant allele.

The duplication of the shorter of these sequences may have arisen as a result of slipped-strand mispairing (Levinson and Gutman, 1987) since there is a region of homology immediately 5′ to this sequence.

The 20 bp insert is not preceded by any regions of notable homology, but may have arisen by a mechanism similar to the type proposed by Kless and Vermass, involving nicking of the nascent molecule, strand displacement, polymerisation of the new strand, and ligation of the displaced and newly formed strands to form the duplication. This is then resolved by replication of the entire molecule.

This system was demonstrated in the case of complementation of deletions in the *Synechocystis* D1 protein of photosystem II by the formation of tandem duplications under nutrient limited conditions (Kless and Vermass, 1995).

The observation that the penultimate residue of the reiterated sequence is different from the 'parental' sequence (G instead of A) is also interesting and may have arisen as a result of an error during replication.

The region including these elements falls outside the range of the structure given by Li *et al.* (Li *et al.*, 2000), so it is not possible to speculate upon the structural implications of these insertions. However it is possible to determine whether hyaluronate lyase activity is affected in this strain. Assays with this mutant (described in the next section) revealed that enzyme activity is unaffected by these insertions, which occur close to the C-terminus of the protein.

3.1.7 – Truncation of the translation product of allele 2

Figure 3.1.5 shows that allele 2 (from strain 12B-13E, ID 922 - a carried serotype 3 isolate) contains an amber stop codon prior to the C-terminal cell wall anchor motif. This is another interesting feature that suggests that the translated protein would no longer be anchored to the cell wall. The strain from which this allele was sequenced belongs to a successful UK clone associated with invasive disease (Muller-Graf, 1999). Sequences of eight other allele 2-type strains were obtained, four from within this clone, and four strains not associated with the clone. The stop codon was found in the *hyl* loci from the clonally related strains, but not the unrelated strains. This will be returned to in Chapter 4, section 4.4.

The hyaluronate lyase phenotype of strain 12B-13E was examined, and is described in section 3.2.8.

Appendix 3.1

Allele	*Tsp*509I	*Mnl*I	*Hinf*I	*Hsp*92II	*Mae*III
1	2	1	1	1	3
2	6	2	1	1	1
3	3	1	1	2	2
4	2	1	1	2	3
5	2	2	1	1	3
6	1	1	1	1	3
7	8	1	1	1	1
8	1	1	1	1	2
9	2	1	1	1	2
10	1	1	1	2	3
11	4	1	1	2	3
12	2	5	1	1	3
13	2	1	1	2	2
14	3	1	1	2	4
15	1	1	2	1	2
16	7	1	1	1	3
17	8	1	1	1	3

Table 3.1.12 - *recP* allele assignments.

Allele	*Alu*I	*Hha*I	*Hinf*I	*Tsp*509I	*Hsp*92II
1	1	1	1	2	1
2	1	1	2	2	2
3	2	1	1	2	2
4	1	1	1	2	2
5	3	1	1	2	2
6	5	1	1	2	3
7	6	1	1	2	2
8	1	2	1	2	2
9	1	1	1	3	2
10	9	1	2	2	2
11	4	1	1	2	2
12	1	1	1	1	3
13	1	1	1	2	3
14	7	1	1	2	2
15	10	1	1	2	2

Table 3.1.13 - *hexA* allele assignments.

Allele	*Dde*I	*Hae*III	*Hinf*1	*Mse*I	*Rsa*I
1	2	1	1	1	1
2	1	1	2	2	1
3	1	2	2	2	1
4	2	1	2	5	1
5	2	1	2	1	1
6	2	1	2	2	1
7	1	1	3	2	1
8	1	1	2	1	1
9	1	1	1	1	1
10	3	1	1	2	1
11	2	3	2	5	1
12	2	1	2	6	1
13	1	6	2	8	1
14	1	1	3	3	1
15	1	4	3	2	1
16	1	1	3	1	1
17	1	1	2	4	1
18	2	1	1	2	1
19	4	5	4	7	4
20	1	2	1	2	1
21	2	1	2	1	2
22	5	7	2	10	3
23	1	1	2	3	1
24	2	3	2	1	1
25	1	1	1	8	1
26	2	1	2	9	1
27	2	1	2	2	5
28	2	1	2	4	1

Table 3.1.14 - *hyl* allele assignments

Carried Strains				Invasive Strains				Serotype Three Strains			
Strain ID	***recP***	***hexA***	***hyl***	**Strain ID**	***recP***	***hexA***	***hyl***	**Strain ID**	***recP***	***hexA***	***hyl***
885	1	1	1	**355**	1	4	8	**470**	10	4	5
922	2	2	2	**53**	14	12	19	**485**	6	2	2
948	3	3	3	**322**	1	4	2	**483**	1	2	2
889	1	4	4	**393**	3	3	2	**481**	1	4	2
914	4	4	5	**39**	3	3	2	**476**	4	4	na
949	1	4	6	**187**	1	5	8	**472**	4	4	na
954	5	5	6	**168**	1	4	8	**471**	4	4	2
956	5	5	2	**169**	1	4	8	**492**	5	2	2
873	6	4	7	**194**	1	5	8	**490**	5	2	26
877	7	5	8	**312**	1	4	8	**489**	5	2	2
965	8	4	8	**339**	1	na	8	**487**	5	2	2
864	9	4	9	**123**	6	5	8	**482**	5	2	2
884	6	4	1	**31**	1	4	16	**480**	5	2	2
902	1	12	10	**91**	1	na	16	**479**	5	2	2
878	1	4	11	**310**	9	4	9	**478**	5	2	2
882	1	4	11	**3**	1	4	8	**477**	5	2	2
903	6	1	5	**324**	15	4	1	**475**	5	2	2
869	6	6	10	**87**	1	4	1	**473**	5	2	27
880	1	7	1	**329**	8	4	1	**466**	5	4	2
862	6	4	8	**303**	8	1	20	**1001**	4	4	2
904	6	4	12	**103**	na	na	na	**474**	5	2	2
872	1	7	8	**326**	1	13	10	**860**	5	2	5
905	10	4	2	**340**	1	14	10	**491**	6	4	2

Carried Strains				Invasive Strains				Serotype Three Strains			
Strain ID	***recP***	***hexA***	***hyl***	**Strain ID**	***recP***	***hexA***	***hyl***	**Strain ID**	***recP***	***hexA***	***hyl***
901	1	4	13	**30**	1	4	2	**494**	6	4	2
906	6	4	8	**59**	1	5	8	**495**	1	2	28
875	1	4	2	**200**	1	4	5	**496**	1	2	28
907	3	4	2	**201**	1	4	1	**497**	1	2	28
863	6	4	14	**213**	1	4	5	**498**	1	2	28
908	1	4	15	**231**	16	4	8	**499**	1	2	28
876	8	5	8	**28**	16	4	5	**501**	1	2	28
871	1	8	16	**237**	16	4	5	**502**	1	2	28
909	11	4	8	**308**	6	4	21	**503**	1	2	28
910	6	4	8	**56**	6	4	6	**504**	1	2	28
911	1	4	1	**319**	6	4	21	**505**	1	2	28
913	8	4	17	**162**	1	4	2				
916	6	5	2	**321**	9	4	8				
917	12	4	2	**75**	11	4	8				
957	12	4	2	**327**	9	4	5				
918	1	4	5	**307**	8	1	22				
952	1	4	na	**26**	17	4	8				
943	1	4	na	**315**	6	4	na				
861	1	4	8	**333**	6	15	23				
919	1	4	17	**161**	6	5	na				
921	1	4	17	**256**	6	4	2				
923	1	4	1	**125**	6	9	8				
866	6	9	2	**318**	5	4	24				
867	1	5	2	**126**	6	4	5				
879	6	10	17	**352**	6	4	5				

Carried Strains				Invasive Strains				Serotype Three Strains			
Strain ID	***recP***	***hexA***	***hyl***	**Strain ID**	***recP***	***hexA***	***hyl***	**Strain ID**	***recP***	***hexA***	***hyl***
945	6	4	18	**379**	1	4	25				
967	13	4	2	**159**	11	4	1				
946	5	4	8	**392**	1	16	18				
947	6	4	17	**145**	4	4	2				
				381	6	4	2				
				398	6	4	2				

Table 3.1.15 – Compiled allele data for all loci studied.

3.2 Assessment of phenotypic variation of hyaluronate lyase production

3.2.1 – Introduction

As mentioned in the introduction (section 1.8.2), the phenotypes of pneumococci producing hyaluronate lyase are believed to vary (Kostyukova *et al.*, 1995). As part of the work addressed in this project, it was decided that the analysis of genetic variation at the *hyl* locus should be complemented by a phenotypic analysis of hyaluronate lyase activity. In order to achieve this, a suitable assay (or assays) for hyaluronate lyase production needed to be found. It is evident from the literature that numerous hyaluronate lyase assays are available, as mentioned in section 1.8.2.1.

It was decided to proceed with the 'Stains-all' assay (to measure any decrease in substrate) alongside the spectroscopic method of Pritchard *et al.* (to monitor the accumulation of digest products). The final decision to proceed with these assays was based on practicability (with the intention to screen a large number of strains a rapid assay was required), cost effectiveness, and technical success in preliminary experiments.

3.2.2 Early experiments with 'Stains-all'

Setting up the Stains-all assay was greatly helped by discussions with Dr. Karen Homer. The assays were performed as described in section 2.2.3. Initially the Stains-all assay was tried using substrate supplied by Sigma-Aldrich Co. (sodium salt of hyaluronic acid derived from bovine trachea). When making aqueous solutions of the material supplied, it was seen that the viscosity of the solution was not high (as expected from the literature), and that the reaction with the 'Stains-all' reagent was also not as expected. Rather than getting a shift from purple to blue colouration, the solution rapidly changed to a straw-yellow colour. This was indicative of some form of bleaching or collapse of the chromophore, similar to the effect of acidifying the 'Stains-all' solution. Attempts to remedy this by making buffered hyaluronate solutions were unsuccessful.

The next step was to switch to an alternative source of hyaluronic acid. The hyaluronic acid derived from human umbilical cord, supplied by ICN Pharmaceuticals, was tried next. This material gave rise to highly viscous aqueous solutions. It also gave the desired colouration upon interaction with 'Stains-all'. This material was used in early experiments to derive standard curves from serial dilutions of hyaluronate solutions using a Dynex spectrophotometer. These curves are shown in figure 3.2.1. It was later noted that the material supplied by Fluka Chemicals was also appropriate for use in these experiments (also from human umbilical cord). It was concluded that differences in the source of hyaluronic acid (and possible differences in preparation at source) were at the source of the early problems encountered using this assay. Attempts to distinguish between the various sources of HA were made by MALDI MS. This indicated that qualitative differences existed between the material supplied by Sigma and those supplied by both ICN and Fluka. These experiments showed that with at least one matrix (dihydroxybenzoate), the distribution of dominant peaks was similar for both the ICN and Fluka material, but different for the Sigma material (data not shown).

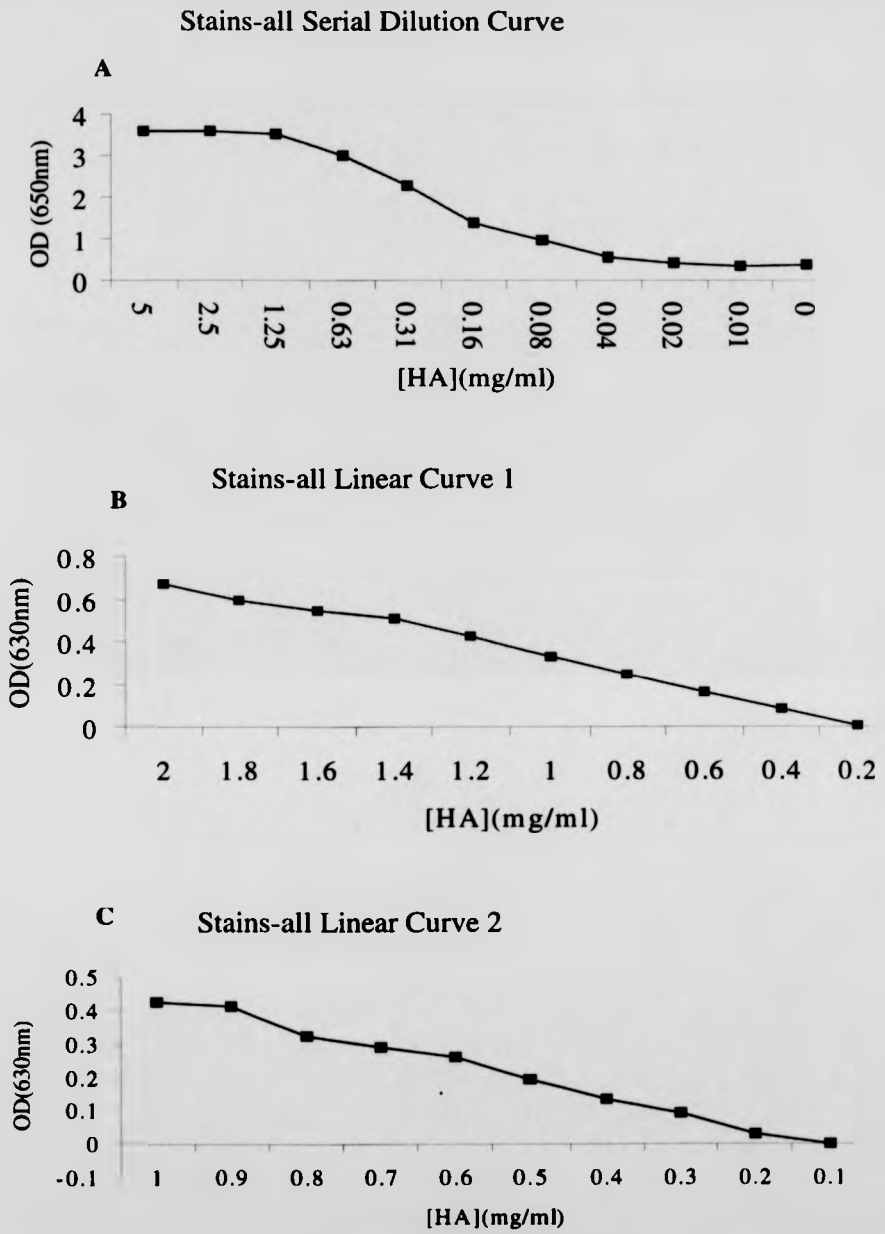
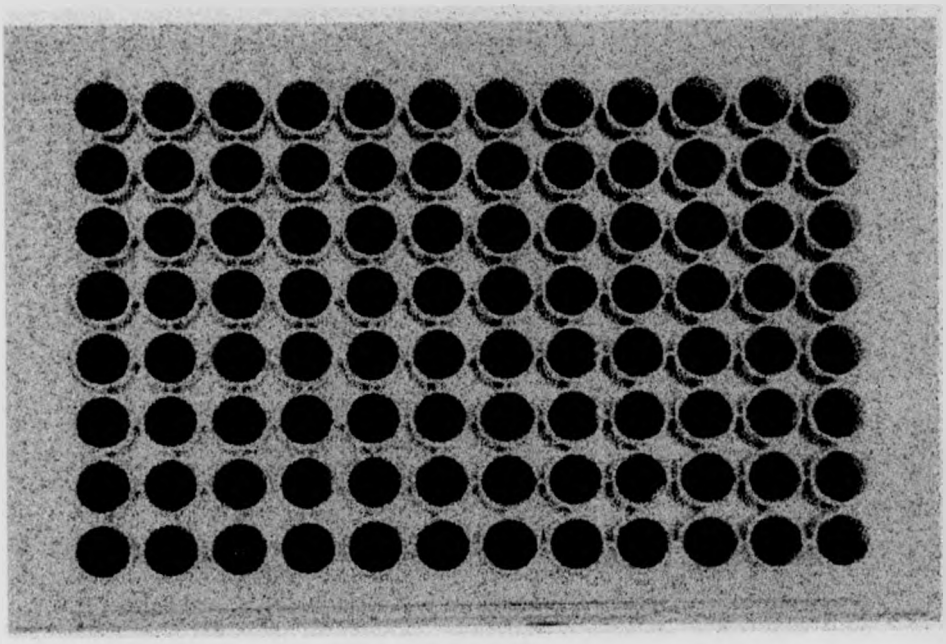


Figure 3.2.1 – Early standard curves obtained using the 'Stains-all' assay. A) Over the range 5 -0mg/ml., B) Over the range 2-0.2mg/ml., and C) over the range 1-0.1mg/ml. Note – The wavelengths used differ by 20nm as these experiments were conducted using different spectrophotometers, equipped with different filters.

Figure 3.2.1 shows the variation of optical density over three concentration ranges. As can be seen, two different wavelengths were used for these experiments (630nm and 650nm). This reflects a change of instrumentation in the lab, the Dynex MRX spectrophotometer (with a 630nm filter) being replaced with the Labsystems iEMS plate reader (with a 650nm filter). Both wavelengths proved to be suitable for the assay. Reasonably good linear responses are seen over the range [2-0mg/ml]. However, it is also apparent that it is not possible to combine the curves. That is, the optical density at 1mg/ml for curve B is not equal to the optical density for the same concentration in curve C. Such variation was repeatedly found during the course of subsequent experiments. It was noted that solutions of 'Stains-all' were not suitable for storage, even at 4°C. Older batches of 'Stains-all' solution gave less coherent standard curves with greater errors between replicates (data not shown). From these observations the following protocols were adopted; for each set of assays, fresh 'Stains-all' solutions were made. For each batch made, standard curves were derived over the concentration range 0.5-0.05mg/ml to ensure that the reagents were performing reliably. An example of such a curve is shown in figure 3.2.2 (below).



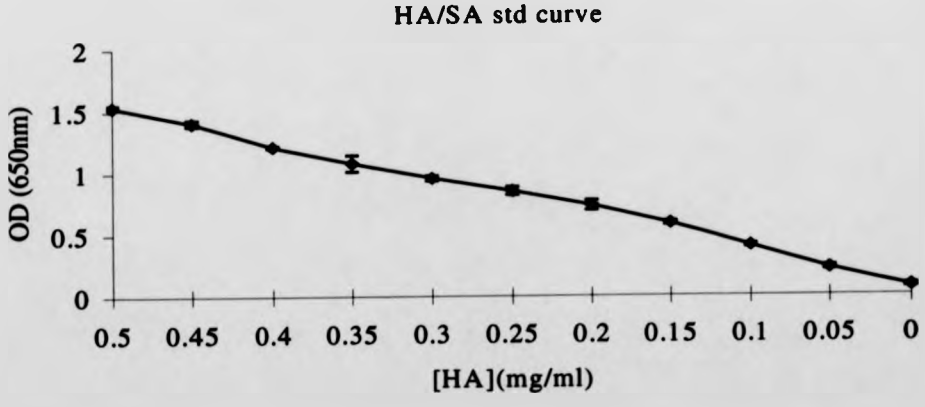


Figure 3.2.2 – Standard curve obtained with the 'Stains-all' assay.

The 'Stains-all' assay gave good linear responses within a particular batch of reagent, but not from batch to batch, making direct comparisons of substrate concentration between sets of experiments difficult. To this end, the assay was conducted in the basis of an internal control comparison. Since the quantity of HA (hyaluronic acid) used in each digest was fixed, negative controls where no cells or crude enzyme preparation were added to the HA were deemed to be comparable between experiments. Any kinetic degradation of substrate over the course of the digest was assumed to be similar over all experiments. Thus the internal negative control reaction gave the maximum expected OD (630nm/650nm) reading for each experiment. The ratio of the OD obtained for a given digest to this maximum value was taken as the basis for the assay.

3.2.3 - The UV/Vis. spectrophotometric assay of Pritchard *et al.*

This spectrophotometric assay described in section 1.8.2.1 proved to be easy to perform on digest mixtures, and was deemed to be a suitable secondary assay to use alongside the 'Stains-all' assay; where one measures the decrease in substrate, the other monitors the accumulation of specific reaction products.

This assay has been used to monitor the accumulation of the disaccharide liberated by hyaluronate lyase (ΔUA GlcNAc) by a number of independent investigators (Lineker and Meyer, 1954, Yamataga *et al.*, 1968, Homer *et al.*, 1994 and Lin *et al.*, 1997).

Control experiments described below also indicated that the assay was a suitable one to use alongside the 'Stains-all' assay previously described. The concentration of the unsaturated N-acetylhyalobiuronic acid product of the reaction is calculated using a millimolar absorption coefficient (ε) of 5.5 (Lin *et al.*, 1997). A standard curve derived using this assay is shown in figure 3.2.2.1.

3.2.4 - Early digests – Induction of hyaluronate lyase

Before hyaluronate lyase assays were conducted on a large range of isolates, a

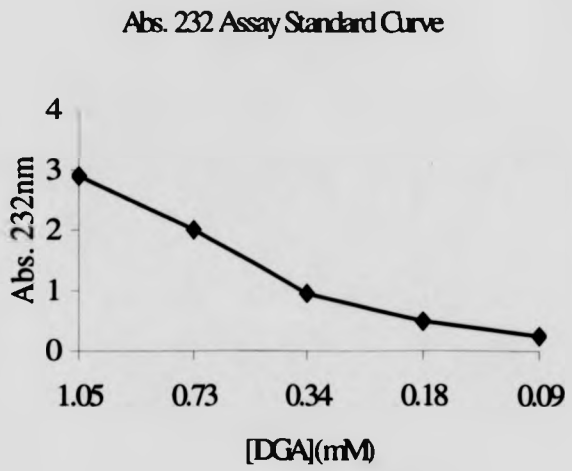


Figure 3.2.2.1 - Standard curve using the Abs. 232nm method. Measurements were obtained from serial dilutions of a hyaluronate digest by strain ID 161 after induction in CMHA.

number of experiments were carried out to qualify the methodology to be employed. These were conducted with a small number of isolates, including the laboratory strain, R6. Digests were set up according to the methodology of Homer *et al.* (*personal communication.*). Briefly, strains were grown overnight on blood-BHI agar plates, and cells harvested into Tris buffer. Cell suspensions were used as crude enzyme preparations on the basis of the LPXTG cell wall anchor motif described in section 1.7.10. The optical densities were adjusted to 0.1 (path-length = 1 cm). This was used as a crude enzyme preparation to inoculate buffered hyaluronate (HA). Degradation of the substrate was seen over time (some examples are shown in figure 3.2.3), as shown by the 'Stains-all' assay (where a reduction in $Abs._{650}$ reflects depolymerisation of the HA substrate). As can be seen, different strains depolymerised the substrate at different rates.

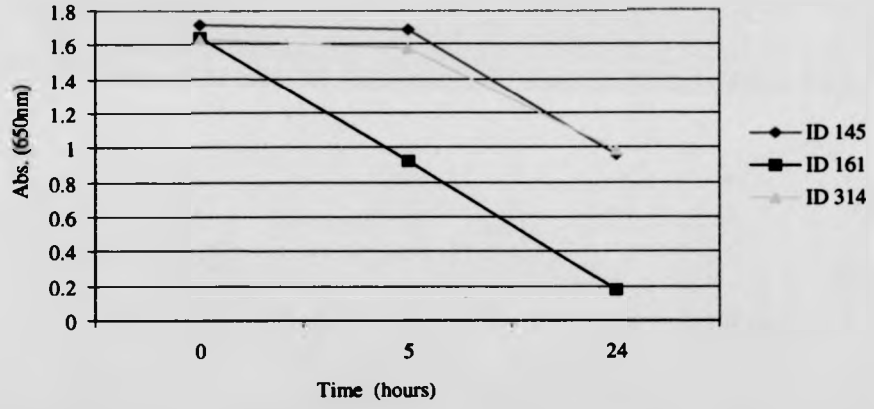


Figure 3.2.3 – Depolymerisation of HA by hyaluronate lyase from pneumococci grown on BHI/blood agar.

The next issue addressed was that of the potential inducibility of hyaluronate lyase production in pneumococci. This has not been reported elsewhere, but such an effect seemed plausible given the nature of the enzyme function. Pneumococci were harvested from BHI-blood agar plates and then incubated in C medium made using no C6 sugars.

To this was added various sugars including glucose, HA, glucuronic acid and N-acetyl glucosamine. These were incubated for periods of time varying from 30 min. to 3 hours. Cells were recovered by centrifugation,

washed and resuspended in Tris buffer as described above. 'Stains-all' assays were conducted, and the results compared with those obtained for those obtained with the same strains taken straight from BHI-agar. The only conditions under which a difference was seen were those in which the cells were incubated in C-medium plus HA (CMHA) for 3 hours (the difference in phenotype was not seen with shorter incubations). An example of such an effect is shown in figure 3.2.4.

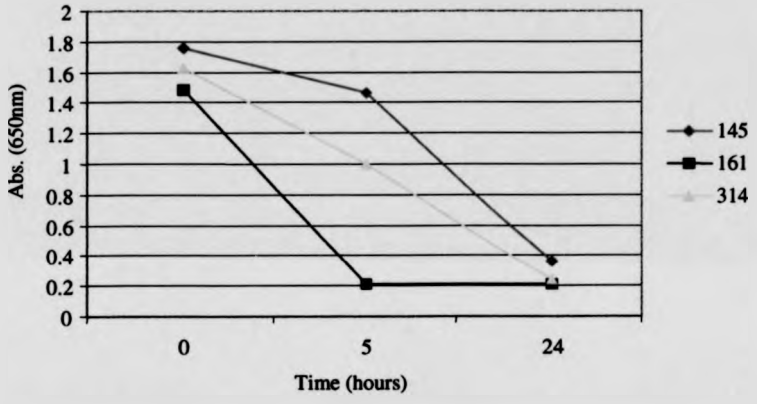


Figure 3.2.4 – Depolymerisation of HA by hyaluronate lyase from pneumococci after induction in CMHA.

As can be seen, for the strains used here, both the rate of depolymerisation, and the overall extent of depolymerisation have been increased by the treatment with CMHA. It was later found that growing the cells on BHI-blood agar, which had been overlaid with aqueous HA could reproduce this effect.

A number of experiments were conducted to determine whether this induction effect was subject to catabolite repression by glucose, by using CMHA containing various concentrations of glucose and N-acetyl glucosamine. In no case was there seen to be a loss of induction in inducible strains.

3.2.5 – Accumulation of hyaluronate lyase activity data

Hyaluronate digests were performed using whole cell suspensions of cells grown overnight on Blood-BHI agar at 37°C. The inducible phenotype of some strains indicated in section 3.2.4 was also further investigated by incubation of aliquots of the same strains in minimal C - medium supplemented with HA (described in detail in section 2.1.1) for 3h prior to inoculation of HA solutions. The 'Stains-all' assays were performed in quadruplicate, and absorbance values of each digest were compared with that of the internal (no hyaluronate lyase) control. The unit of activity was defined as a reduction of the internal (no hyaluronate lyase) control value by 10%, after Benchetrit (Benchetrit *et al.*, 1978). The mean absorbance value and standard deviation was calculated for each set of readings. Units of activity were calculated on the basis of the mean values obtained, with 95% confidence limits (CL) based on the standard deviation of each set of quadruplicate readings.

In some cases, the $Abs._{650}$ for the sample was slightly higher than that for the negative control. For ease of computation these values were recorded as activity 'not detected'. The higher absorption in these assays indicated that there was more HA in the sample after the digest than before. Since no HA synthetase has been reported in pneumococci, this was thought to be unlikely. The increase in $Abs._{650}$ has been attributed to one of two possible factors; The presumed presence of other complex polysaccharides in the solution, transferred with the whole cell fraction. It is arguable that if this is the case, carryover of interfering substances (which react with 'Stains-all' in the same way as HA) may mask hyaluronate lyase activity. Alternatively, the higher OD's may simply reflect the variability of this assay system, and form part of the inherent error in the measurements.

The aim of these studies was to determine whether variation in hyaluronate lyase activity amongst a population of strains correlates with either site of isolation (disease state), or allelic variation at the *hyl* locus. Since some strains give a negative result with the 'Stains-all' assay, it is arguable that these results cannot be taken to be true measures of activity. However, since two assays have been used, it is suggested that independent treatment of these data sets will indicate correlations

if there are any. That is to say, where a statistically significant result is missed using one assay, it is picked up in the second.

The tables in appendix 3.2 (tables 3.2.5 – 3.2.9) show the data obtained for the strains studied. They are divided into five groups representing the state from which each strain was isolated. These are blood (bacteraemia), carried (asymptomatic), lung (lower respiratory tract infection, or pneumonia), CSF (meningitis) and ear (otitis media). Activities under non-inducing ('Normal') and inducing conditions are given.

3.2.6 – Statistical treatment of hyaluronate lyase activity data

In the first instance the average values of each set of titres, for each group were compared. These values are shown in table 3.2.1.

	Normal SA Average Titre	Induced SA Average Titre	Normal [ΔUA GlcNAc]	Induced [ΔUA GlcNAc]
Blood	4.425	7.052	0.42	0.79
Carried	3.004	6.379	0.36	0.79
Lung	1.575	3.917	0.12	0.46
CSF	3.501	6.701	0.31	0.73
Ear	3.445	6.095	0.25	0.65

Table 3.2.1 – Average hyaluronate lyase titres of the strains included in this study.

By observation, several points can be drawn from the data in this table: it is apparent that the activities of the groups of isolates are similar - with the exception of the lung isolates, which have lower average titres. Also it seems that induction of hyaluronate lyase occurred in all groups. However, the data analysed here are average values, taken from groups with high variability between individual strains. It was considered necessary to formalise the analysis of these observations.

When looking for variation between population samples such as these, the statistical method of 'analysis of variance' or anova is used. This technique compares the ratio of variation within a group (in this case groups are blood, CSF

etc.), with variation between the groups (blood vs. CSF, vs. carried etc.), and is described in some detail in chapter 2, section 2.2.4.

Since the numbers of strains included in each group were not equal, a model II anova was used, which weights the intra group variance in accordance with the group sample size.

The output of the anova procedure (F_s) is compared with a predetermined value (F_{crit}), derived from a theoretical probability distribution. This ratio of F_s to F_{crit} is then used to either prove or disprove the null-hypothesis, which states that two populations vary in the same way, and therefore do not differ significantly from one another. The results of these calculations are given in table 3.2.2 below.

	Stains-all				**[ΔUA GlcNAc] Assay**			
	F_s	F_{crit}	P	N-H?	F_s	F_{crit}	P	N-H?
Normal	2.862	2.421	0.024	No	3.302	2.424	0.012	No
Induced	3.193	2.421	0.004	No	2.554	2.424	0.040	No

Table 3.2.2 – Results of F tests over all groups of isolates studied. N-H – is the null-hypothesis supported?

As can be seen, in all cases the null hypothesis has been rejected. That is to say $F_s > F_{crit}$, and $P<0.05$. Thus in both the normal, and the induced data sets, using both assays, there is an added element (or elements) of variation, indicating a significant statistical difference between the groups.

In order to establish the source of this variation, the Tukey-Kramer method was used to make unplanned post-hoc comparisons of the data. This method is described in section 2.2.4.1. The results of this analysis show that it is the lung isolates that lead to the increased variance from the means during the Anova procedure. The activity of hyaluronate lyase in this group of strains is lower than that of the carried, CSF and blood isolates, but not of the otitis isolates. Combinations of other pairs of groups do not lead to significant differences beyond that expected by chance, and repetition of the Anova without the lung strains included leads to the acceptance of the null hypothesis in all cases. The data obtained from the Tukey-Kramer procedure are given in table 3.2.3.

'Stains-all' Assay					
	Blood	Carried	CSF	Lung	Otits
Blood	X	2.299 2.511	2.191 2.395	2.519 2.750	2.905 3.175
Carried	1.420 0.673	X	1.608 1.759	2.029 2.218	2.492 2.727
CSF	0.924 0.303	0.496 0.370	X	1.909 2.087	2.395 2.619
Lung	**2.850** **3.136**	1.430 2.463	**1.926** 2.833	X	2.696 2.947
Otitis	0.980 0.957	0.440 0.284	0.056 0.654	1.870 2.179	X
[ΔUA GlcNAc] Assay					
	Blood	Carried	CSF	Lung	Otits
Blood	X	0.231 0.358	0.219 0.316	0.262 0.370	0.293 0.146
Carried	0.059 0.008	X	0.162 0.235	0.212 0.304	0.250 0.354
CSF	0.114 0.069	0.055 0.077	X	0.204 0.289	0.243 0.347
Lung	**0.304** 0.329	**0.245** 0.337	0.190 0.260	X	0.277 0.397
Otitis	0.171 0.140	0.112 0.148	0.057 0.071	0.133 0.189	X

Table 3.2.3 – Output of the Tukey-Kramer post-hoc analysis of the Anova results. The upper right portion of each matrix gives MSD, the lower left gives actual difference between means. Blue denotes data obtained under non inducing conditions, red denotes induced activity data. Values that differ significantly are underlined.

Importantly, there is no evidence presented to support the findings of Kostoyukova *et al.* who associated high hyaluronate lyase titres with CSF isolates (Kostoyukova *et al.*, 1995).

This work does seem to indicate that there is an induction effect upon pre-incubation of isolates in media containing HA. This indication was investigated using the paired T-test. This was carried out for data from both assays and is presented in table 3.2.4 below.

	Stains-all				[ΔUA GlcNAc] Assay			
	T_s	T_{crit}	P	N-H?	T_s	T_{crit}	P	N-H?
Normal vs Induced	14.705	2.131	6.22×10^{-5}	No	23.681	2.131	9.43×10^{-6}	No

Table 3.2.4 - Paired T-test results investigating the significance of the induction effect on mean titres.

These results clearly show that there is a statistically significant effect of incubation of isolates in media containing HA prior to assay, by two independent assay data sets.

3.2.7 – Can hyaluronate lyase phenotype be related to genotype?

The question of correlating phenotypic observations with the allele designations defined in the previous chapter was addressed next. The groups of isolates used for the allelic variation assessment and the phenotypic assessment were not identical, however there was a large group of isolates (64 in total) from the phenotypic study whose *hyl* alleles had been defined by HRRA. Once again, anova was used to investigate the variance within the individual groups (based on allele type this time) and across the groups. The null hypothesis in this case was 'there is no correlation between hyaluronate lyase titre and representative allele'. Since the anova procedure requires that groups contain more than one value (some allele types were represented by only one strain), the total number of strains used in this study was reduced to 54. The alleles represented were 1, 2, 5, 8, 10, 11, 16 and 17 (alleles 1, 2, 5, 8, and 17 were found to be the most common in the work outlined in section 3.1). As before, data sets from both assays were used. Data taken from assays of hyaluronate lyase activity under normal and inducing conditions were compared also. These are shown in table 3.2.4.1.

	Non-induced		Induced	
Allele	'Stains-all'	[ΔUA GlcNAc]	'Stains-all'	[ΔUA GlcNAc]
1	3.749	0.46	7.735	0.91
2	3.028	0.32	6.191	0.80
5	3.238	0.35	8.434	0.91
8	3.885	0.35	6.766	0.73
10	5.272	0.75	4.729	0.78
11	2.475	0.12	0.655	0.17
16	1.260	0.11	5.901	0.43
17	0.037	0.06	0.711	0.14

Table 3.2.4.1 – Average hyaluronate lyase titres of allele representative strains.

As before, one way Anova was performed on the data to assess the extent of phenotypic variation between the strains studied. Using the criteria outlined in the

previous section, the null hypothesis was accepted for the non-induced data set ('Stains-all' assay: F_s=0.818, F_{crit}=2.216, P=0.576, [ΔUA GlcNAc] Assay: F_s=1.066, F_{crit}=2.221, P=0.4). The situation for the induced data set was complicated by the rejection of the null hypothesis for the 'Stains-all' assay (F_s=3.064, F_{crit}=2.216, P=0.0098), and the acceptance of the null hypothesis using the data from the [ΔUA GlcNAc] Assay (F_s=1.593, F_{crit}=2.221, P=0.161).

The rejection of the null hypothesis for the 'Stains-all' data was investigated using the Tukey-Kramer method as described previously. The results of this show that the strains bearing *hyl* alleles 11 and 17 showed less hyaluronate lyase activity than those bearing allele 1, and that the allele 11 strains were also significantly less active than the allele 8 strains. The sample sizes of the allele 11 and 17 strain groups were low however (n = 2 and 3 respectively), and so other representatives of these alleles would need to be analysed before any firm conclusions can be drawn from this result.

3.2.8 – Activity of the truncated hyaluronate lyase encoded by allele 2

As mentioned in section 3.1.7, allele 2 contains a stop codon in the hyaluronate lyase ORF at position 835 with respect to the 949 amino acid protein. The cell wall anchor motif lies C-terminal to this, so it was hypothesised that if an active protein were to be produced by this allele, activity might be found in the cell culture supernatant, rather than associated with the cell.

The strain from which allele 2 was sequenced (ID 922) was grown in C-medium until early exponential phase. Cells were removed by centrifugation and the pellet resuspended in tris buffer. Culture supernatant and resuspended cells were used in HA digests as previously described, and the production of ΔUA GlcNAc was measured at 232nm after incubation at 37°C for 24 hours. The results of this experiment are shown in figure 3.2.5, below.

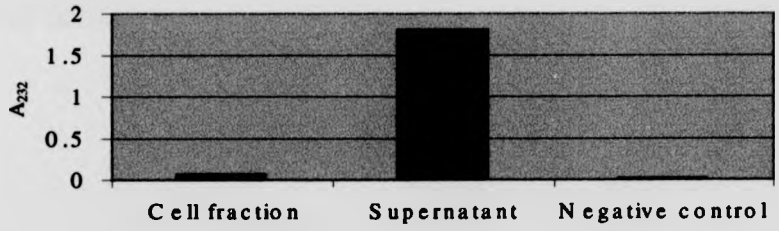


Figure 3.2.5 – Production of ΔUA GlcNAc by hyaluronate lyase encoded by allele 2 from strain ID 922.

As can be seen from figure 3.2.5, hyaluronate lyase activity is high in the supernatant fraction but negligible activity is associated with the cell. This is a clear indication that allele 2 from strain 922 is able to produce an active enzyme, which dissociates from the cell. It is unlikely that supernatant activity is seen due to autolysis of the growing cells, as the experiment was carried out using pneumococci growing at early exponential phase.

This observation runs counter to the assertion of Berry *et al.* (Berry *et al.*, 1994) that hyaluronate lyase is released from the cytoplasm in a LytA dependent manner, and supports the hypothesis outlined in chapter 1 - section 1.8.6, that the *hyl* ORF as described is incomplete,

and is likely to include a gram positive signal sequence. This question is returned to in section 3.6.

Appendix 3.2

Hyaluronate lyase activity data

The final column in each table indicates whether or not the given strain is inducible The total number of inducible strains is shown at the bottom of this column as a proportion of those strains studied in each group. N(C) denotes that induction was not seen as the strain produced high hyaluronate lyase titres under non- inducing conditions also. ΔUA GlcNAc is given as the abbreviation of the pneumococcal hyaluronate lyase reaction product, and concentration is expressed in mM. This crude value was taken as the titre in these assays. NA indicates that data were not available. Av denotes the average value of activities given. nd =activity not detected. Induction was scored as an increase in titre (using either assay) of ≥ 20%.

Blood	Normal			Induced			Induction?
Strain	Units (SA)	95%CL	[ΔUA GlcNAc]	Units (SA)	95%CL	[ΔUA GlcNAc]	
36	nd	-	0.06	2.463	0.725	0.12	Y
31	1.472	0.761	0.11	7.144	0.106	0.48	Y
75	1.530	1.508	0.05	1.231	0.762	0.06	N
77	0.627	0.957	0.13	9.099	0.084	1.15	Y
81	8.154	0.077	0.56	9.064	0.046	1.14	Y
87	9.270	0.055	1.02	9.071	0.089	1.00	N(C)
91	2.310	2.302	0.11	2.870	0.551	0.15	N
107	nd	-	0.11	0.424	1.259	0.09	Y
110	9.280	0.045	0.82	9.117	0.029	1.14	N(C)
116	8.806	0.071	0.71	9.117	0.029	1.09	Y
117	0.528	0.385	0.10	3.037	0.794	0.20	Y
145	8.074	0.414	0.58	9.120	0.042	1.11	Y
161	9.155	0.058	1.14	9.103	0.029	1.09	N(C)
314	8.621	0.231	0.67	9.099	0.021	1.31	Y
324	6.832	0.035	0.44	9.099	0.038	1.07	Y
329	1.142	0.908	0.07	8.849	0.244	0.66	Y
342	1.373	1.053	0.29	9.078	0.028	1.02	Y
353	nd	-	0.13	6.261	0.591	0.35	Y
354	4.614	3.116	0.58	8.762	0.164	1.79	Y
355	6.723	0.317	0.78	9.037	0.073	0.89	Y
Av	4.425		0.42	7.052		0.79	15/20

Table 3.2.5 – Hyaluronate lyase activities of strains isolated from bacteraemic patients.

Carried	Normal			Induced			Induction?
Strain	Units (SA)	95%CL	[ΔUA GlcNAc]	Units (SA)	95%CL	[ΔUA GlcNAc]	
864	nd	-	0.09	4.317	0.832	0.45	Y
885	nd	-	0.23	8.927	0.588	0.57	Y
914	4.536	0.241	0.49	8.641	0.616	1.08	Y
949	nd	-	0.11	nd	-	0.35	Y
965	0.814	0.426	0.31	8.951	0.570	1.38	Y
905	3.802	0.257	0.36	9.039	0.054	1.44	Y
901	2.544	1.517	0.51	5.793	0.074	0.51	Y
880	2.461	0.790	0.37	4.733	0.110	0.41	Y
862	5.242	0.725	0.52	9.053	0.055	1.37	Y
872	2.738	0.578	0.15	4.391	0.085	0.60	Y
878	2.022	0.934	0.16	1.310	0.089	0.28	N
882	2.929	1.015	0.09	nd	-	0.07	N
903	2.430	1.030	0.07	9.070	0.048	0.70	Y
869	1.292	0.465	0.29	0.219	0.151	0.23	N
865	nd	-	0.06	7.378	0.039	0.58	Y
922	1.263	0.624	0.36	1.066	0.144	0.60	N
884	9.141	0.062	1.39	9.265	0.044	2.02	N(C)
902	9.252	0.073	1.22	9.239	0.043	1.34	N(C)
873	8.390	0.076	0.64	9.141	0.046	0.96	Y
877	7.216	0.190	0.60	9.222	0.039	1.35	Y
904	2.022	0.324	0.22	9.248	0.038	1.33	Y
956	8.382	0.115	0.68	9.193	0.042	1.11	Y
948	8.294	0.121	0.78	9.248	0.034	1.24	Y
945	nd	-	0.10	8.966	0.437	0.73	Y
943	nd	-	0.05	4.815	0.859	0.45	Y
867	8.894	0.109	1.05	8.974	0.393	1.32	N(C)
879	8.863	0.090	1.09	8.936	0.441	1.43	N(C)
923	nd	-	0.08	8.931	0.455	1.33	Y
866	2.504	0.705	0.33	7.026	0.520	0.81	Y
919	nd	-	0.06	0.665	1.026	0.17	N
921	nd	-	0.06	1.468	0.915	0.21	Y
861	nd	-	0.11	8.927	0.468	0.34	Y
947	nd	-	0.07	2.524	0.624	0.35	Y
909	8.137	0.588	1.19	8.940	0.072	1.22	N(C)
908	0.664	1.260	0.12	8.858	0.119	0.98	Y
906	8.292	0.543	0.78	8.970	0.111	1.08	N(C)
917	2.535	0.634	0.26	8.974	0.105	1.25	Y
918	8.912	0.451	1.11	8.927	0.072	1.26	N(C)
967	1.088	1.369	0.13	6.386	0.168	0.62	Y
876	4.288	0.656	0.34	4.597	0.305	0.39	N
952	0.686	1.345	0.10	8.927	0.079	1.07	Y
913	0.111	1.521	0.05	nd	-	0.05	N

916	0.049	1.266	0.12	8.893	0.085	1.15	Y
863	3.894	0.547	0.45	8.884	0.116	1.14	Y
871	nd	-	0.12	7.691	0.307	0.67	Y
875	0.044	2.074	0.07	1.871	0.693	0.26	Y
907	0.783	2.496	0.13	7.107	0.176	0.73	Y
910	1.584	1.212	0.11	1.871	0.975	0.19	N
911	1.150	0.690	0.09	3.009	0.658	0.22	Y
Av	3.004		0.36	6.379		0.79	34/49

Table 3.2.6 – Hyaluronate lyase activities of strains isolated from asymptomatic carriers.

Lung	Normal			Induced			Induction?
Strain	Units (SA)	95%CL	[ΔUA GlcNAc]	Units (SA)	95%CL	[ΔUA GlcNAc]	
3	nd	-	0.05	1.233	0.674	0.19	Y
28	0.291	1.345	0.09	9.167	0.530	0.94	Y
162	nd	-	0.03	nd	-	0.03	N
166	0.029	0.774	0.09	nd	-	0.01	N
167	nd	-	NA	0.011	0.193	NA	N
307	0.459	1.007	0.07	0.552	1.181	0.11	Y
432	nd	-	0.05	1.967	1.492	0.21	Y
304	7.328	0.144	0.03	8.972	0.181	1.18	Y
123	5.449	0.218	0.39	7.673	0.098	0.50	Y
154	5.859	0.230	0.44	9.113	0.056	1.22	Y
176	1.404	0.316	0.21	9.130	0.055	1.07	Y
301	6.632	0.489	0.51	8.870	0.069	1.07	Y
303	nd	-	0.04	4.676	0.100	0.34	Y
308	2.969	0.536	0.27	9.141	0.058	1.22	Y
312	1.026	0.703	0.05	9.059	0.059	1.23	Y
315	0.078	0.547	0.09	2.666	0.072	0.18	Y
317	1.333	0.925	0.03	2.038	0.098	0.26	Y
339	nd	-	NA	8.831	0.052	NA	Y
467	nd	-	0.04	1.094	0.068	0.05	Y
468	0.403	0.759	0.05	0.232	0.169	0.03	N
26	9.315	0.047	0.09	9.379	0.384	0.70	N(C)
27	nd	-	NA	nd	-	NA	N
44	nd	-	0.02	nd	-	0.03	N
45	0.522	0.356	0.24	9.311	0.313	0.96	Y
46	0.753	1.256	0.03	nd	-	0.03	N
103	0.504	1.738	0.03	0.477	1.106	0.02	N
106	0.771	0.857	0.02	nd	-	0.01	N
139	0.298	1.078	0.03	nd	-	0.07	N
163	0.265	1.512	NA	nd	-	NA	N

Av	1.575		0.12	3.917		0.46	17/29

Table 3.2.7 – Hyaluronate lyase activities of strains isolated from lower respiratory tract infections.

CSF	Normal			Induced			Induction?
Strain	Units (SA)	95%CL	[ΔUA GlcNAc]	Units (SA)	95%CL	[ΔUA GlcNAc]	
1795	9.244	0.035	1.36	8.993	0.544	1.03	N(C)
1796	3.510	0.746	0.16	9.440	2.456	1.12	Y
1797	3.324	0.320	0.21	7.199	0.851	0.64	Y
1798	5.308	0.409	0.36	0.034	1.367	0.09	N
1799	1.722	0.288	0.15	7.913	0.777	0.61	Y
1800	6.110	0.502	0.83	9.450	1.758	0.62	Y
1801	2.487	0.684	0.36	8.916	0.555	1.01	Y
1802	2.344	0.127	0.10	8.318	0.757	0.53	Y
1803	0.253	0.532	0.09	8.903	0.608	1.01	Y
1804	4.912	0.226	0.20	8.959	0.585	0.95	Y
1805	nd	-	0.05	8.924	0.469	0.98	Y
1806	2.192	1.234	0.19	8.933	0.589	1.05	Y
1807	2.636	0.424	0.14	8.924	0.654	1.06	Y
1808	2.539	1.702	0.18	8.924	0.607	1.25	Y
1809	2.477	0.557	0.08	5.968	0.799	0.40	Y
1810	nd	-	0.02	nd	-	0.09	N
1811	0.749	0.438	0.17	7.461	0.726	0.67	Y
1812	nd	-	0.01	nd	-	0.07	N
1813	9.303	0.062	1.14	8.963	0.618	1.19	N(C)
1814	5.927	0.433	0.10	7.341	0.629	0.36	Y
1815	0.062	0.219	0.14	9.023	0.635	1.16	Y
1816	nd	-	0.11	8.804	0.704	0.78	Y
1817	7.276	0.291	0.33	7.672	1.374	0.66	N(C)
1818	1.086	0.469	0.05	6.829	0.704	0.68	Y
1819	1.638	0.480	0.03	9.840	1.534	1.02	Y
1820	8.907	0.093	0.82	8.920	0.664	1.02	N(C)
1821	5.290	0.342	0.36	8.723	0.720	0.72	Y
1822	1.383	1.430	0.24	8.924	0.689	0.88	Y
1823	8.466	0.060	0.74	8.949	0.670	1.09	N(C)
1824	7.087	0.160	0.53	8.961	0.666	1.10	Y
1825	9.034	0.056	0.92	9.020	0.606	1.05	N(C)
1826	0.409	0.551	0.14	5.226	0.517	0.46	Y
1827	7.162	0.209	0.57	9.003	0.721	1.13	Y
1828	7.719	0.277	0.62	8.936	0.694	1.09	Y
1829	1.204	0.602	0.24	nd	-	0.13	N
1830	2.576	0.404	0.28	2.835	1.235	0.39	N

1831	7.532	0.150	0.58	9.038	0.591	1.10	Y
1832	6.807	0.266	0.30	9.170	0.498	1.26	Y
1833	7.279	0.453	0.51	8.405	0.640	0.65	Y
1834	5.888	0.165	0.26	9.119	0.573	1.25	Y
1835	1.472	0.821	0.11	9.140	0.540	0.78	Y
1836	6.869	0.202	0.49	9.071	0.580	0.90	Y
1837	8.301	0.271	1.06	9.056	0.637	1.25	N
1838	1.999	1.398	0.11	9.100	0.573	0.68	Y
1839	2.739	0.522	0.09	8.603	0.524	0.38	Y
1840	7.686	0.060	0.43	9.067	0.610	0.91	Y
1841	nd	-	0.47	1.898	0.674	0.30	N
1842	1.985	0.643	0.06	3.770	0.898	0.26	Y
1843	nd	-	0.08	5.256	0.677	0.41	Y
1844	nd	-	0.16	8.774	0.652	0.98	Y
1845	nd	-	0.11	8.765	0.672	1.04	Y
1846	nd	-	0.15	7.283	0.766	0.53	Y
1847	nd	-	0.04	nd	-	0.05	N
1848	nd	-	0.06	nd	-	0.04	N
1849	nd	-	NA	nd	-	NA	N
1850	nd	-	0.29	8.216	0.901	0.95	Y
1851	nd	-	0.07	2.731	1.005	0.28	Y
1852	nd	-	0.03	nd	-	0.04	N
1853	nd	-	0.35	8.788	0.671	1.00	Y
1854	nd	-	0.11	6.958	1.063	1.19	Y
1856	8.887	0.072	0.63	9.383	0.337	0.90	Y
1857	2.251	1.025	0.50	3.589	0.611	0.80	Y
1860	3.763	1.068	0.90	3.511	0.905	1.00	N
1861	2.927	0.197	0.04	2.518	1.263	0.06	N
1862	3.685	0.569	0.06	3.865	0.419	0.05	N
1863	3.608	0.828	0.05	3.228	0.851	0.06	N
1864	3.967	0.225	NA	3.083	1.433	NA	N
1865	4.646	0.598	NA	4.395	0.816	NA	N
1866	5.387	0.169	NA	9.386	0.328	NA	Y
1867	9.430	0.039	NA	9.400	0.298	NA	N
1868	7.162	0.243	NA	9.412	0.298	NA	Y
Av	3.501		0.31	6.701		0.73	47/71

Table 3.2.8 – Hyaluronate lyase activities of strains isolated from CSF.

Ear	Normal			Induced			Induction?
Strain	Units (SA)	95%CL	[ΔUA GlcNAc]	Units (SA)	95%CL	[ΔUA GlcNAc]	
63	1.900	0.959	0.05	9.215	0.369	0.77	Y
59	3.705	0.695	0.09	3.658	0.409	0.17	N
197	4.388	0.523	0.14	9.321	0.380	1.10	Y
192	3.834	0.611	0.09	9.305	0.412	0.92	Y
10	3.693	0.557	0.10	9.297	0.403	0.96	Y
65	5.098	0.275	0.21	9.302	0.441	1.86	Y
55	3.780	0.269	0.04	3.343	1.015	0.10	N
347	5.207	0.243	0.17	8.892	0.378	0.52	Y
492	2.610	1.148	0.06	nd	-	0.04	N
481	2.366	1.110	0.32	9.305	0.395	0.76	Y
425	nd	-	0.06	nd	-	0.07	N
430	9.261	0.056	1.20	9.231	0.398	1.20	N(C)
428	9.285	0.099	1.20	9.288	0.399	1.40	N
213	nd	-	0.23	6.367	0.571	0.52	Y
474	nd	-	0.03	nd	-	0.02	N
473	nd	-	0.05	1.000	0.278	0.07	N
Av	3.445		0.25	6.095		0.65	8/16

Table 3.2.9 – Hyaluronate lyase activities of strains isolated from the middle ear.

3.3 – Hyaluronate lyase negative mutants – Natural and experimental

3.3.1 – Introduction

The work described in this section concerns the functional deletion of *hyl*, and phenotypic characterisation of the resultant mutants. Two mutants are described. One is a naturally occurring deletion mutant and one has been generated by insertion-recombination mutagenesis. These are described in turn.

3.3.2 - Restriction mapping the deletion from strain 619 (ID 53)

As mentioned in section 3.1.3.3, one of the isolates (619 - strain ID 53) examined gave rise to a *hyl* PCR product that was approximately 400 bp smaller than the wild type product. This strain was a vaginal sepsis isolate. The small PCR product was either the result of spurious amplification of a non-specific product (deemed unlikely due to the relatively high annealing temperature used in the PCR reaction, and the absence of this sized product from any other strain), or the loss of part of the *hyl* gene. Data from the HRRA analysis described in section 3.1 were used to map this proposed deletion. Figure 3.3.1 is a representation of the various digests used in the HRRA procedure. The presence or absence of bands representing the respective digest products is indicated.

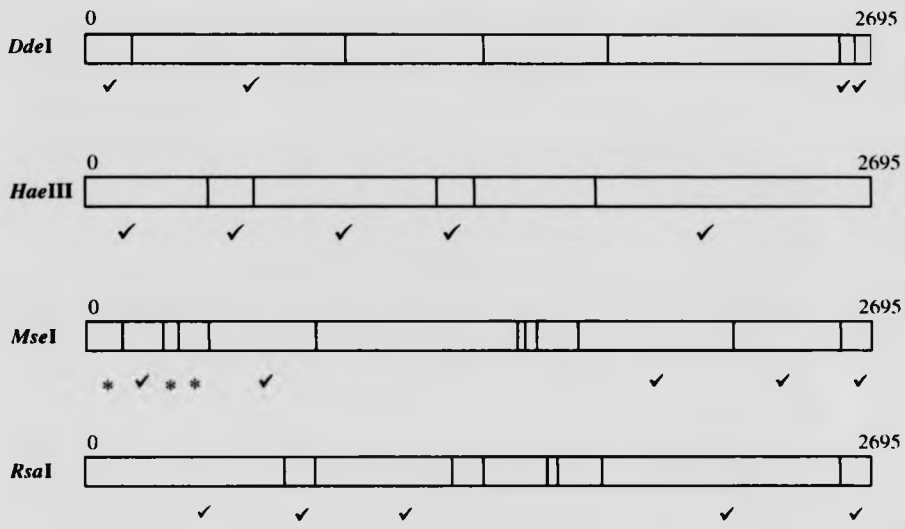


Figure 3.3.1 - Diagrammatic representation of the expected restriction sites on the database *hyl* sequence (approximate positions shown). Ticks below each fragment indicate that the particular band was seen with the deletion mutant allele 19. Asterisks indicate that the band was expected but not seen. This was because the expected size of the band was too small or the quality of the stained gel was not good enough to visualise the fragment. The *Hind*III pattern is not shown as the high numbers of restriction sites complicate the analysis, and do not substantially contribute to the definition of the deletion described below.

The most informative of the restriction patterns seen with this allele arose from *Hae*III and *Rsa*I digests. The absence of the expected bands between positions1178 bp to 1720 bp (with respect to the PCR product of the database gene sequence) with the *Rsa*I digest, and positions 1283 bp to 1690 bp with the *Hae*III digest indicate that the deletion has occurred in this region.

Analysis of the *Mse*I digest supports this conclusion; a band migrating at about 470 bp is unique to allele 19. The presence of this band can be explained by looking at the *Mse*I sites in the 829 bp to 1698 bp region, shown in figure 3.3.2 (below).

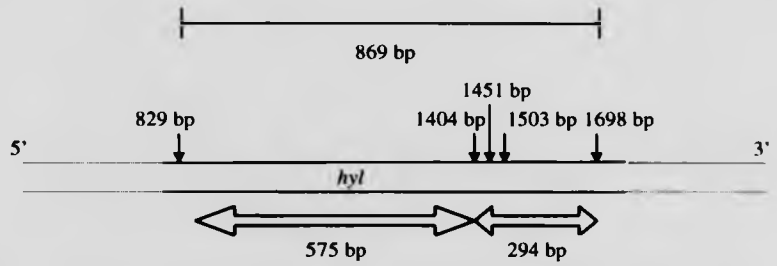


Figure 3.3.2 - Representation of the region around 800-1700 bp of the *hyl* sequence. Arrows indicate *Mse*I sites, with positions shown in base pairs with respect to the *hyl* PCR product.

The overall size of the deletion is estimated to be 400 bp, and the 3′ end of the deletion is arbitrarily placed at the 3′ most extremity of the estimated range of the deletion (ie; position 1697 bp on the *hyl* PCR product). It can be seen that we loose the 294 bp encompassing the three *Mse*I sites 5′ to this. We also effectively shorten the 575 bp fragment by 106 bp (≈ 400 bp – 294 bp). This would give rise to a fragment of 469 bp, which is in very close agreement with the 470 bp band observed.

The extent of variation in the 3′ direction was not thought to be large, as if this were the case, the 176 bp *Hae*III band (which is observed) would be noticeably smaller on the gel, due to the deletion encroaching on this region. The overall likely error in this approximation, taking into account estimating the size of the deletion, and variation of the 176 bp *Hae*III band was estimated to be not more than 20-30 bp. Thus the site of the deletion was described as between nucleotides 1298 and 1698 with respect to the *hyl* PCR product. These values correspond to nucleotides 1513-1913 on L20670, the *hyl* sequence described by Berry *et al.* (Berry *et al.*, 1994).

Subsequent sequence analysis showed this estimation to be accurate. Using the predicted site of the deletion described above, the appropriate sequencing primer (hsp.R2) was used to sequence over the junction of the deletion. This corresponded

to 407bp, between nucleotides 1277 and 1684 of the *hyl* PCR product, or 1492 and 1899 of L20670.

3.3.3 - Phenotypic analysis of the naturally occurring deletion mutant

Strain 619 was phenotypically Hyl$^-$ using both assays employed in this study. Also, this strain was unable to utilise HA as a sole source of carbon (see section 3.4).

As mentioned in section 3.1.2, Southern analysis of PCR negative strains revealed that *hyl* is present in the chromosome as a single copy. This confirms the observation of Berry *et al.* (Berry *et al.*, 1994) using a larger number of strains. However, this does not rule out functional degeneracy, where hyaluronic acid may be degraded by a factor genetically distinct from hyaluronate lyase.

The mutation seen in *hyl* from strain 53 was seen to arise from a sporadic deletion event. It is not known how long progenitors of this strain might have contained this deletion, nor what other mutations might have accumulated within its genome. With this in mind it was decided that strain 53 may not provide a reliable example of a *hyl* negative mutant. To address this, a defined *hyl* mutant was generated by the insertion of a chloramphenicol acetyl transferase (CAT) cassette into the *hyl* ORF.

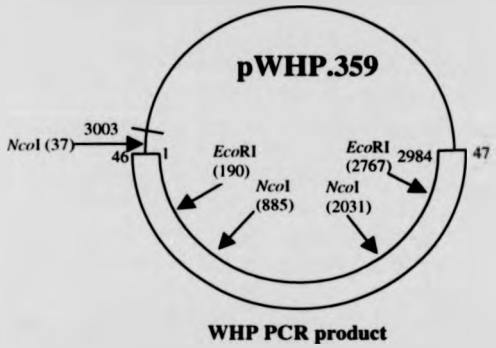
3.3.4 – Directed mutation of *hyl* by insertional inactivation

3.3.4.1 – Generation of a WHP::CAT plasmid construct

As mentioned above, the phenotype of strain 53 was tentatively attributed to the loss of part of the *hyl* ORF. The potential for functional redundancy could not be reliably commented on, as the genetic background of this strain is unknown. In order to generate an isogenic knockout the *hyl* ORF was disrupted by the insertion of the chloramphenicol acetyl transferase (CAT) cassette.

The molecular genetic approaches employed are outlined in chapter 2, and the rationale initially adopted is described in figure 3.3.3, below.

Step 1 involved cloning the WHP PCR product from strain R6 (ID 359) into a pUC18 derivative prepared for TA cloning (see chapter 2, section 2.2.1). This was analysed by restriction analysis to confirm the insertion and orientation of the WHP product within the vector. This is indicated below.



*Eco*RI	*Nco*I
3410	3947
2577	1146
	894

Expected sizes of digest products based on the structure of pWHP.359 are given in the table. Results of screening 6 putative transformants by this restriction analysis are shown below.

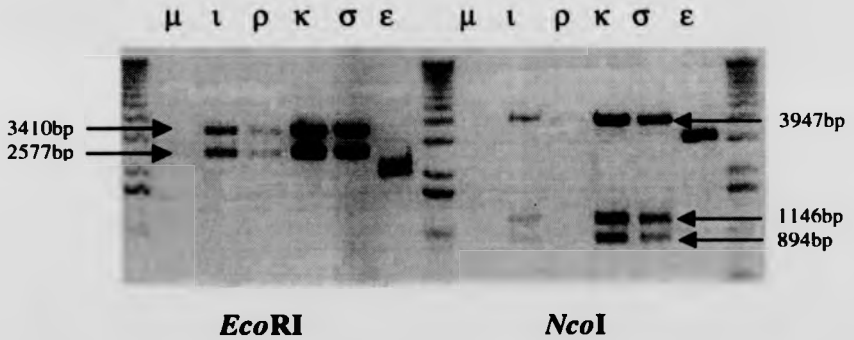


Figure 3.3.3 – Screening minipreps of putative whp clones. Top: Diagramatic representation of the expected clone, with the whp product inserted in the 'forward' direction with respect to the vector sequence. Positions marked are given with respect to the vector (outside the circular portion) and the insert (inside the circular portion). Expected restriction sites are assigned according to the supplied vector sequence and the available *hyl* database sequence (accession no. L20670). Bottom: Restriction digests showing the products generated with the 6 putative recombinants screened. The markers used were kb ladder.

As can be seen, 4 of the minipreps tested gave the expected sized products on digestion. One of these (κ) was kept and named pWHP.359.

The central portion of the cloned WHP product was removed by digestion with *Eco*RI (see figure 3.3.3). The fragment containing the vector backbone, plus the whp 5' and 3' regions was purified from 1% agarose after electrophoresis. A 901bp *Eco*RI CAT fragment (encoding the chloramphenicol acetyl transferase expression cassette - kindly donated by Sue Colby) was ligated to the pWHP.359 fragment. Transformations were carried out as described in chapter 2

Putative recombinants were isolated after expression and growth on the appropriate selective media. The WHP, CAT and TA primer pairs were used in various combinations in order to screen the putative recombinants and determine the orientation of CAT with respect to the WHP sequence. This is depicted in figure 3.3.4.

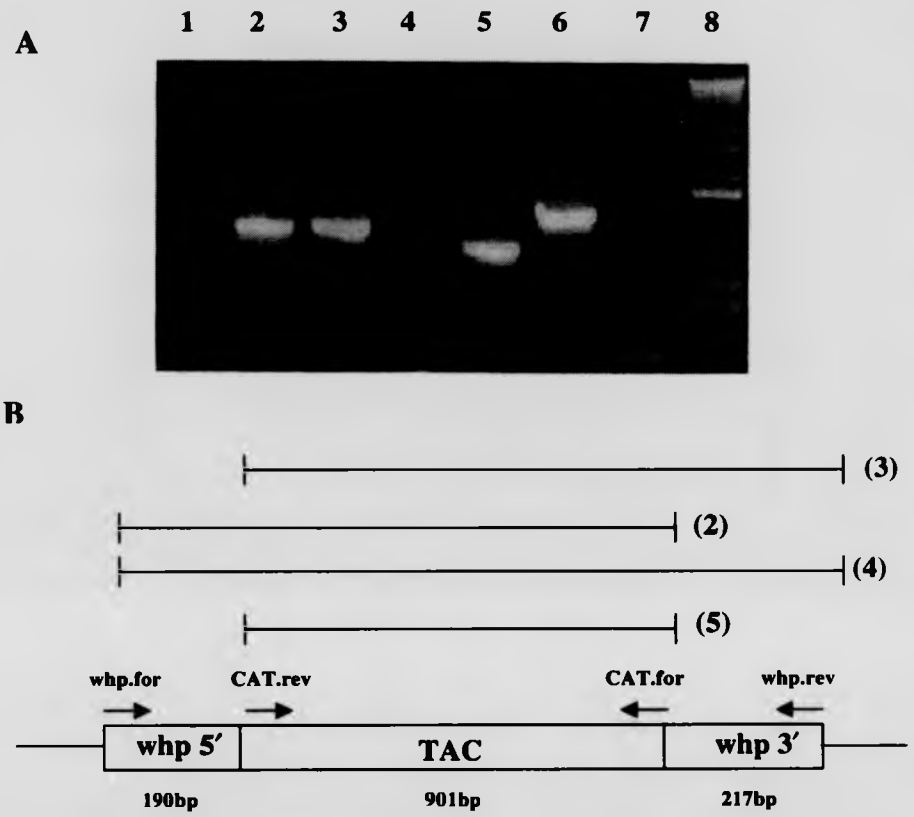


Figure 3.3.4 – A: PCR confirmation of the location and orientation of the CAT cassette within the whp cloned hyaluronate lyase ORF. Lane designations; 1- whp.for/CAT.rev, 2 – whp.for/CAT.for, 3- whp.rev/CAT.rev, 4- whp.rev/CAT.for, 5- CAT.for/rev control, 6- whp.for/rev control, 7- no DNA control with all primers, lane 8 - Kbl. B: Diagrammatic representation of the resultant recombinant, termed pWHP.359::CAT. The extent of the PCR products giving the bands seen on the gel are indicated by horizontal lines, with the respective lanes on the gel indicated in brackets.

This construct was used to transform of the lab strain R6. Selection was carried out on Blood-BHI agar plus chloramphenicol. Despite multiple rounds of transformation, no transformants were obtained with this construct. One possible

explanation for this lies in the observation that the regions of homology between the construct and the chromosomal sequence are separated by approximately 2.5Kb with respect to the wild type sequence. The sites which mediate potential recombination are relatively small (around 200bp) with the overall size of the wild type hyaluronate lyase gene.

3.3.4.2 – Generation of a second *hyl::CAT* construct

The next approach was slightly different. As there were no convenient staggered restriction sites in either pWHP.359 or the CAT *Eco*RI fragment a TA cloning approach was taken. This procedure exploits the terminal transferase activity of *Taq* polymerase, which preferentially adds dATP to the 3′ hydroxyl end of a double stranded PCR product. In the absence of other nucleotides however, other residues can be added. The A-tagged CAT PCR product was obtained by simple PCR amplification of CAT from p683eCAT1 (a pneumolysin::CAT construct donated by Sue Colby), using the CATfor/rev primer pair. This was purified after agarose gel electrophoresis.

The vector was digested with *Bst*1107I site (aka *Bst*Z17I - blunt cutting) which cleaves once at position 1833 bp with respect to the *hyl* numbering in figure 3.3.3. This was purified after agarose gel electrophoresis, and T-tagged as described in chapter 2, section 2.2.1. The PCR product using the CAT primers was ligated into this prepared vector, and transformed into *E. coli*.

7 putative transformants were found which were screened by PCR using the TA primers and hyl-2° primers. One particular putative transformant (γ_C) gave rise to products of the expected size using both primer pairs and was further characterised using the TA, hyl-2°, and CAT primers in various combinations. The results of this experiment are shown in figure 3.3.5.

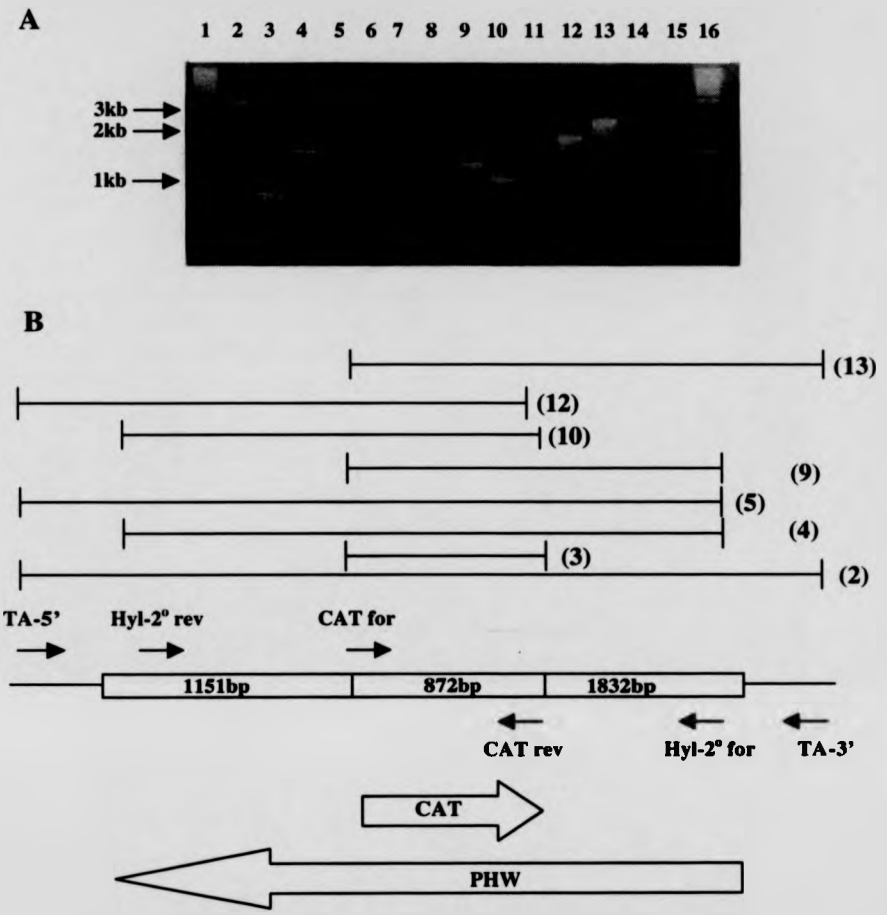


Figure 3.3.5 – Generation of a second recombinant *hyl::CAT* construct. A; The putative recombinant plasmid (γ_C) was screened using the TA, hyl-2° and CAT primers in various combinations. Lane designations; 1 - Kbl, lane 2 - TA-5′/3′, lane 3 - CAT for/rev, lane 4 - hyl-2°for/rev, lane 5 - TA-5′/hyl-2° for, lane 6 - TA-3′ hyl-2° rev, lane 7 - hyl-2° for/CAT rev, lane 8 - hyl-2° rev/CAT for, lane 9 - hyl-2° for/CAT for, lane 10 - hyl-2° rev/CAT rev, lane 11 - TA-5′/CAT for, lane 12 - TA-5′/CAT rev, lane 13 - TA-3′/CAT for, lane 14 - TA-3′/CAT rev, lane 15- no DNA control, all primers, lane 16 - Kbl.
B; The results of the PCR analysis are interpreted as shown. The diagram is not to scale, and the relative positions of the primer annealing sites have been omitted for simplicity. The extent of the PCR products giving the bands seen on the gel are indicated by horizontal lines, with the respective lanes on the gel indicated in brackets. The TA vector backbone is depicted as a single line, the cloned *hyl* and

CAT sequences as boxes. The large arrows at the bottom of the diagram indicate the direction of the two ORF's with respect to the 5'-3' orientation of the TA vector cloning site.

The structure depicted in figure 3.3.5B is supported on the basis of the theoretically calculated sizes of PCR products using the various primer combinations, and the estimated sizes of the products shown in figure 3.3.5 - A. This putative clone was kept and renamed pWHP::CAT2.

This was used to transform strain R6 using CSP1. After two rounds of transformation, followed by screening of putative crm resistant colonies by PCR using the CAT and hyl-2° primer pairs, a single putative recombinant, P41, was found. This was screened by PCR in a similar way to the plasmid construct, pWHP::CAT2, the results of which are shown in figure 3.3.6.

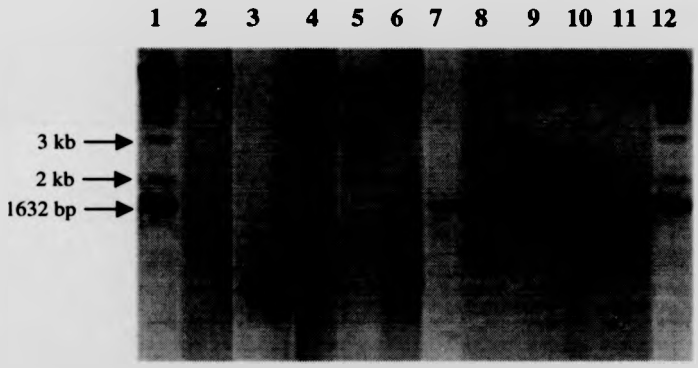


Figure 3.3.6 – Photomontage showing the PCR analysis of the recombinant strain p41. Lane designations; 1 – Kbl, 2 - R6 vs CAT for/rev, 3 - P41 vs CAT for/rev, 4 - R6 vs whp for/rev, 5 - P41 vs whp for/rev, 6 - R6 vs hyl-2° for/rev, 7 - p41 vs hyl-2° for/rev, 8 - P41 vs whp for/CAT for, 9 - P41 vs whp rev/CAT rev, 10 - P41 vs hyl-2° for/CAT for, 11 - P41 vs hyl-2° rev/CAT rev, 12 - Kbl.

The original gel image has been rearranged for ease of presentation. Lanes 1 and 12 contain kb ladder (kbl). The first 6 lanes between these (lanes 2-7) are arranged in pairwise groups, showing the results of PCR with the CAT, whp and hyl-2° primer pairs. Each pair is arranged as wild type (R6) then P41. As can be seen, the

CAT element amplifies from P41, but not the parental wild type strain R6. There is an increase in the sizes of the products obtained using the whp and hyl-2° primer amplified from strain p41 relative to the parental strain R6.This size increase represents the additional 872bp of the CAT cassette.

The orientation of CAT within *hyl* was confirmed by the reactions shown in lanes 8-11. These reactions include one of the CAT primers with either the whp or hyl-2° primers used to confirm the orientation of the CAT insert in the plasmid pWHP359::CAT (see above). These confirm that the CAT insert is in the reverse orientation with respect to the *hyl* ORF. The reciprocal negative controls were also performed, which did not lead to the amplification of any products. In the interests of brevity, these reactions are omitted from figure 3.3.6.

The results of this experiment also reveal that the knockout has arisen as a result of a double crossover recombination event, not a single crossover. The latter event would lead to the incorporation of the entire length of the plasmid element into the chromosomal *hyl* sequence. This would also mean that in some of the CAT/whp or hyl-2° PCR's, two products would be obtained. One arising from the CAT primer vs whp/hyl-2° anealing within the cloned *hyl* derived sequence, plus on from the CAT primer vs whp/hyl-2° anealing with the chromosomal copy of *hyl*. As can be seen from the gel image in figure 3.3.6, this is not the case.

A further confirmation of the structure of the *hyl* knockout was obtained by restriction analysis. The two CAT-*hyl* PCR products (those obtained with the whp primers) were digested with the restriction enzyme *Sty*I. This enzyme was selected on the basis of the predicted cleavage sites shown from the available sequence data for *hyl* and *cat* (accession numbers L20670 and J01754 respectively). This experiment is described in figure 3.3.7.

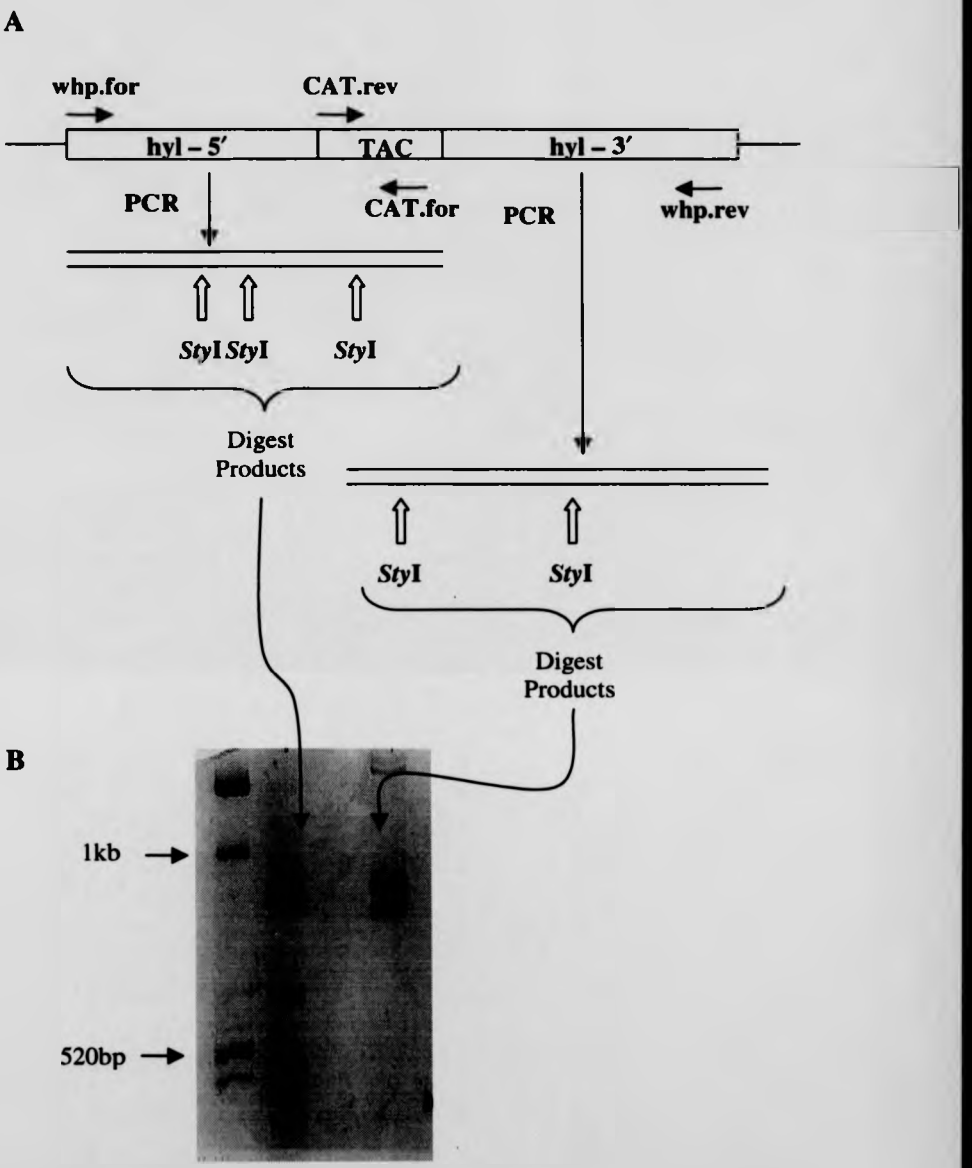


Figure 3.3.7 – Confirmation of the *hyl* interposon knockout structure by restriction analysis. A; Diagramatic depiction of the theoretical structure of the region, showing PCR products generated using the whp and CAT primers (described in the text above). *Sty*I restriction sites are approximated. B; Separation

of the digest products through 8% acrylamide, visualised using ethidium bromide. Markers = Kbl.

The sizes of the restriction digest products for the *hyl* 5′ plus *CAT* PCR product are 909, 860, 600 and 336 bp. Those for the *hyl* 3′ plus *CAT* PCR product are 929, 821 and 272 bp. These were visualised by PAGE, shown in figure 3.3.7B. It was concluded that strain P41 was the desired hyaluronate lyase knockout strain. This was confirmed by sequencing across the junctions of the *hyl::CAT* insertion (data not shown).

3.3.5 – Phenotypic characterisation of experimental *hyl* mutant P41

The aim of this mutagenesis was to address the question of functional degeneracy of hyaluronate lyase in *S. pneumoniae*. Put another way, can a structurally unrelated protein reproduce the depolymerisation of hyaluronic acid in the absence of a functional *hyl* product?

To answer this question, interposon mutagenesis was conducted with the hyl^+ lab strain, R6, generating the $hyl^-:crm^R$ mutant, P41. The insertion site of the CAT cassette leads to the disruption of the *hyl* ORF. The inserted sequence (which lies in the reverse orientation with respect to the *hyl* ORF) results in stop codons in all three open reading frames. Thus read-through of the hyaluronate lyase transcript is blocked.

Strain P41 was used in hyaluronate lyase assays (described in chapter 3.2), along with the parental strain R6. Both the Stains-all and ΔUA-GlcNAc assays were used.

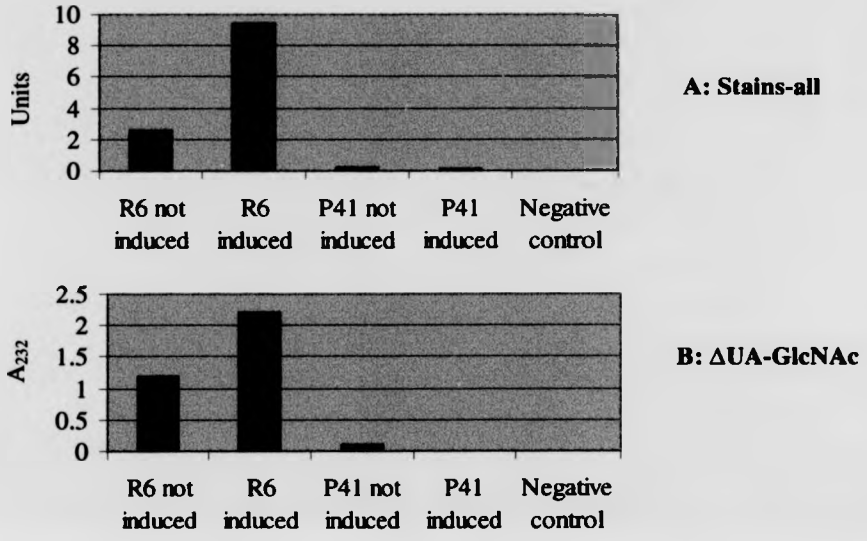


Figure 3.3.8 – Hyaluronate lyase assays using the hyaluronate lyase mutant P41 and R6. A; Stains all assay. B; ΔUA-GlcNAc assay.

As can be seen, using both assays available, the mutant strain P41 is unable to degrade hyaluronic acid. This clearly shows that there is no functional redundancy with respect to hyaluronic acid degradation by the pneumococcus.

The next section describes the ability of some pneumococci to utilise hyaluronic acid as a sole carbon source, and the effect that this mutation has on growth.

3.4 – Pneumococcal growth on hyaluronate

3.4.1 – Introduction

This section describes experiments conducted to demonstrate the ability of some pneumococcal strains to utilise hyaluronic acid as a sole carbon and energy source. Pneumococci cultured *in vitro* are supplied with readily metabolisable sources of carbon, in the form of complex carbohydrates and glucose in BHI medium, and glucose and sucrose in C-medium. The situation *in vivo* is quite different. For instance, in the case of pneumococcal carriage, strains have the ability to persist in the nasopharynx subsisting on the available sources of carbon.

As mentioned in section 1.7.1, hyaluronic acid is estimated to comprise as much as 50% of the human extracellular matrix (ECM) (Frost *et al.*, 1996). It is suggested here that aside from the function of hyaluronate lyase as a means of tissue invasion, it may play a role in the acquisition of metabolisable carbon from host hyaluronic acid. In this manner, hyaluronate lyase may contribute to the ability of the pneumococcus to persist in the human host.

3.4.2 – Formulation of defined media

The fist stage of these experiments was to generate media restrictive for pneumococcal growth, to which chosen carbon sources could be added. It was felt that BHI was by nature too complex to work with, in that the precise composition of the medium is unknown. The composition of pneumococcal C-medium on the other hand, is well defined. C-medium was formulated as described in chapter 2, section 2.1.1, omitting the C6 sugars glucose and sucrose. Experiments with a number of strains showed that this medium was unable to support pneumococcal growth, an example of which is shown in figure 3.1.1, using strain GM 108 (ID 154), a Spanish serotype 3 strain isolated from the lower respiratory tract.

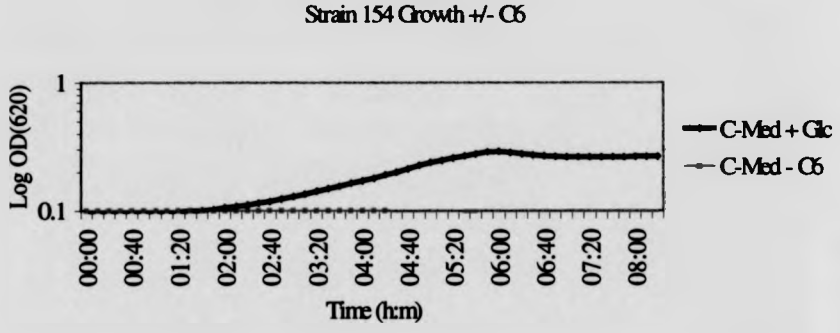


Figure 3.4.1 – Growth of strain 154 in C-medium without glucose or sucrose, and with C-medium plus glucose at 890 μM. 890 μM was defined as the glucose concentration that supports limited growth of pneumococci as determined by R. Waite (unpublished data – *personal communication*).

As can be seen, strain 154 was unable to grow on C-medium formulated without glucose or sucrose. This indicated that strain 154 was unable to subsist on other components of C-medium (such as amino acids or yeast extract). This observation was repeated with all strains studied. Thus a model system was defined which could be easily used to study the ability of pneumococci to grow on alternative carbon sources.

3.4.3 – Growth using HA as a sole source of carbon

The next series of experiments addressed the ability of hyaluronate to support growth of pneumococci. This experiment was performed with a large number of strains using C-medium that contained no added glucose or sucrose, to which aqueous HA at 10mg ml^{-1} had been added to give a final concentration of 2mg ml^{-1}. The results of such an experiment are shown in figure 3.4.2 below. The strains used are described in table 3.4.1.

Strain	Serotype	Site of Isolation	Year of Isolation
359 (R6)	nc*	n/k	1940's
471	3	n/k	1955
1001 (D39)	3	n/k	1996
53	1	Vagina	1987 (Spain)
322	3	n/k	1991 (Kenya)
355	1	Blood	1995 (Uruguay)
312	7	Sputum	1989 (Kenya)
39	4	CSF/Blood	1988 (Spain)
393	4	n/k	n/k
194	6B	Eye	1989 (Spain)
169	6B	Wound infection	1989 (Spain)

Table 3.4.1 – Strains used in the growth experiment shown in figure 3.4.2. *nc – non-capsular. n/k – not known.

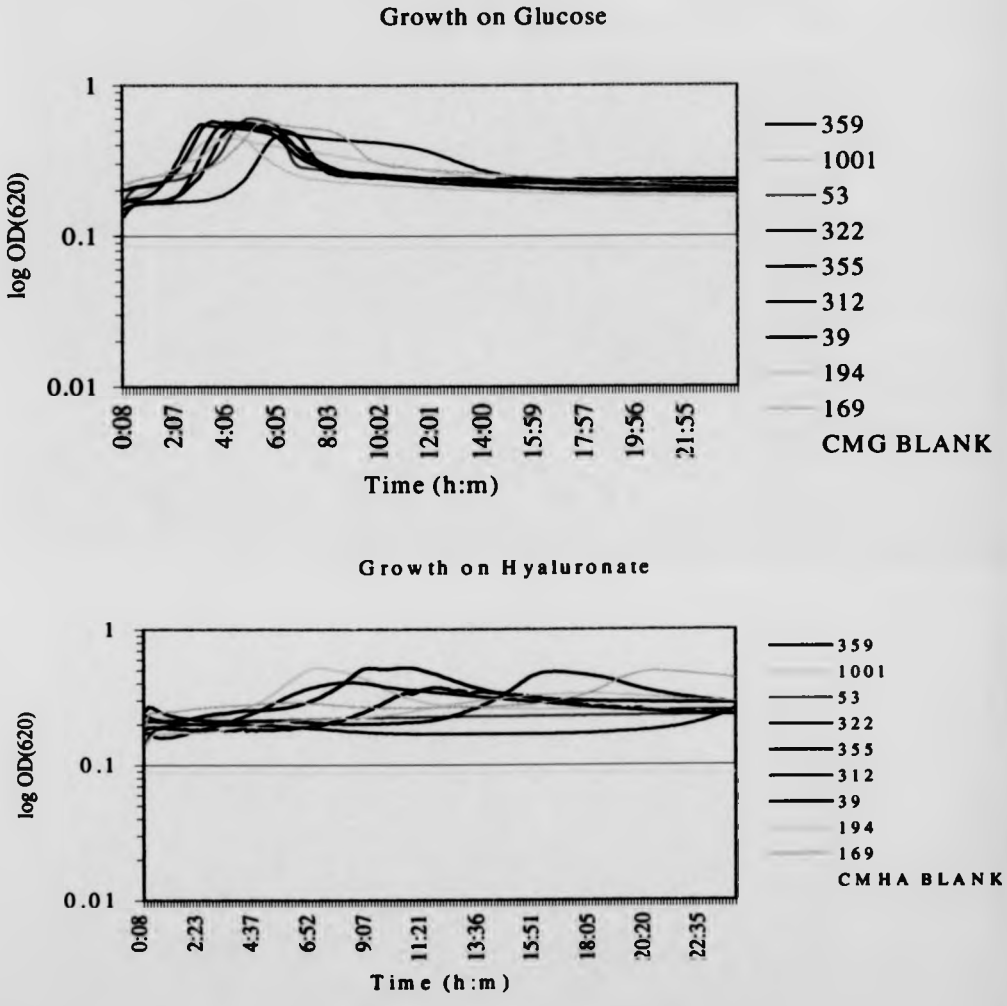


Figure 3.4.2 – Growth of a range of strains on glucose and hyaluronate.

As can be seen, hyaluronate was able to support limited growth of some strains. Another feature of the experiment is the range of phenotypes seen with respect to growth on hyaluronate. Some strains were able to grow readily on HA with a short

lag period (e.g.; strain 1001), whereas some did not begin to grow until significantly later (e.g.; strain 355, where O.D(620) began to increase approximately 11 hours after the start of the experiment). In all, 26 of the 36 strains studied (72%) showed some growth using HA as a sole carbon source. One possible suggestion for the occurrence of growth late in the time course is that components of dead and lysed cells are released, enabling the saprotrophic growth of a sub-population of cells. This is not thought to be the case, as experiments carried out using C-medium without added carbon failed to produce growth late in the time course with any strain.

This experiment is similar to one carried out using *Streptococcus intermedius* by Homer *et al.*, which showed that a particular strain (UNS35) could also utilize hyaluronic acid (Homer *et al.*, 1997). They were also able to identify an intracellular glycuronidase, responsible for cleavage of the hyaluronate lyase disaccharide product.

In the case of *S. intermedius*, growth on HA was shown to be catabolite repressed by glucose. Catabolite repression of hyaluronate lyase expression is not thought to occur in *S. pneumoniae* since 'Stains-all' assays of the culture supernatant showed that HA was still degraded when strains were cultured in C-medium plus HA that contained limiting amounts of glucose (890μM) (R. Waite – personal communication).

3.4.4 – Hyaluronate lyase is required for growth on HA

It was assumed that the first step in the process of hyaluronate utilization is the breakdown of the polymer into disaccharides by hyaluronate lyase. This suggestion is supported by the observation that strain 619 (ID 53 - the deletion mutant described in section 3.3.2) is unable to grow on hyaluronate (see figure 3.4.2, above). It was decided that this assumption would be more vigorously tested by looking at the growth phenotype of the lab mutant P41, described in section 3.3.4. The results of this experiment are shown in figure 3.4.3, below.

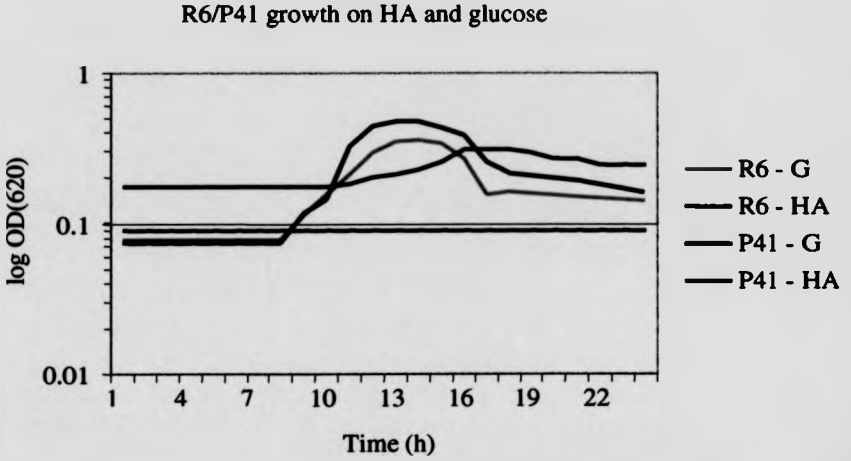


Figure 3.4.3 – Growth of the hyaluronate lyase mutant P41, and its parental strain R6 on C-medium supplemented with glucose or HA.

As can be seen from figure 3.4.3, both R6 and P41 grow well on C-medium plus glucose, but only R6 is able to grow on C-medium plus HA. Thus it is concluded that functional hyaluronate lyase is required for the utilization of hyaluronic acid as a sole carbon and energy source. It is not known whether P41 is able to grow on disaccharide alone.

3.4.5 – Further investigation of growth on HA derived metabolites

This section was concluded by a brief investigation of possible events in pneumococcal growth on HA downstream of hyaluronate lyase activity. It has been shown that growth requires hyaluronate lyase, which presumably performs the initial degradative reaction. It is further assumed that carbon utilization requires subsequent liberation of the N-acetylglucosamine and glucuronic acid derived moieties, from the ΔUA-GlcNAc disaccharide product. This step would involve

another enzymatic cleavage. A further requirement is the uptake of either the disaccharide product of hyaluronate lyase, or its constituent monosaccharides described above.

C-medium without glucose or sucrose was supplemented with either N-acetylglucosamine or glucuronic acid (both at 8.7mM), and tested for the ability to support growth in the strains used in the experiment shown in figure 3.4.4.

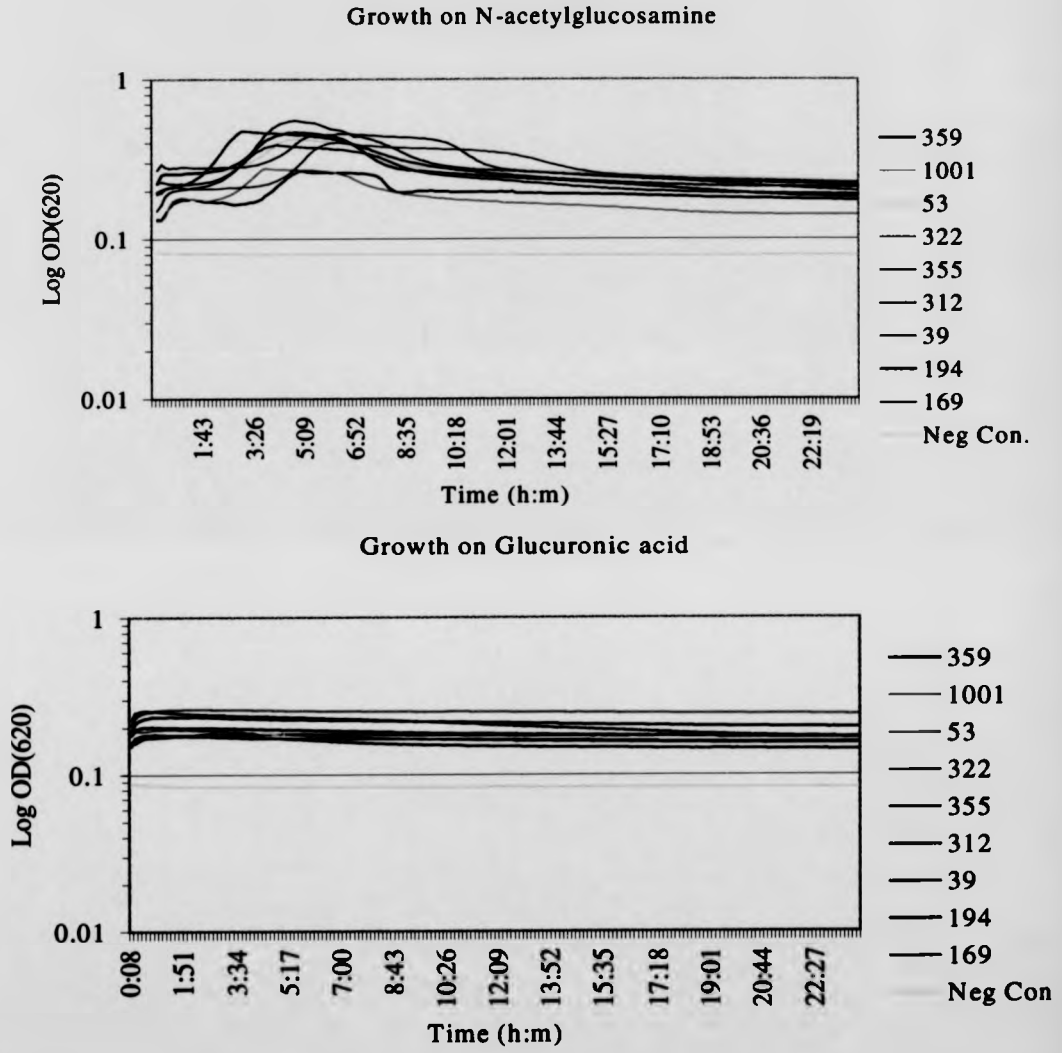


Figure 3.4.4 – Growth of a range of strains on N-acetylglucosamine and glucuronic acid.

The growth curves in figure 3.4.4 show that all strains grow well on N-acetylglucosamine as a sole carbon source, but not at all on glucuronic acid. Note also that strain 619 (ID 53), which was shown to be phenotypically Hyl$^-$ in section 3.3.2 was also able to grow on GlcNAc. Glucuronic acid failed to support growth with any strain.

3.4.6 – Discussion

The data presented above indicate that *S. pneumoniae* is able to utilize hyaluronic acid as a sole carbon and energy source (supplemented with other growth factors in defined media which are unable to support growth themselves). Further, growth on hyaluronate has been shown to be dependent on hyaluronate lyase.

N-acetylglucosamine supported growth with all strains tested, including the *hyl* deletion mutant, strain 619 (ID 53). A suggested pathway for GlcNAc utilisation is offered in Chapter 4, section 4.4.

The case of glucuronic acid is complicated by the fact that the residue found in hyaluronate is the β–4,5 unsaturated derivative. If the liberation of GlcNAc from ΔUA-GlcNAc (assumed to proceed via a glycuronidase activity as found by Homer *et al.*) results in the glucuronic acid derivative becoming fully saturated, then the experiment above is sufficient to demonstrate that glucuronic acid derived from hyaluronate is not a suitable substrate for pneumococcal growth. However, if the residue remains in the unsaturated form, it is not possible to speculate on its ability to contribute to carbon utilization in pneumococci on the basis of experimental evidence presented above. Another possibility is that the organism is not exposed to exogenous glucruronic acid as the disaccharide product of the hyaluronate lyase reaction is transported into the cell prior to cleavage by an intracellular glycuronidase. This is speculated upon in chapter 4, section 4.5.

These observations support the hypothesis that pneumococcal hyaluronate lyase contributes to virulence not only in its role as a spreading factor, but also enables the pneumococcus to persist in the host by metabolising host derived polysaccharides. The biochemical basis of HA metabolism is speculated upon in chapter 4.

3. 5 – The effect of the capsule on hyaluronate lyase phenotype

3.5.1 – Introduction

The classic experiments of Avery, MacCleod and McCarty identified DNA as the genetic material responsible for mediating the transformation of *S. pneumoniae* between rough (acapsular) and smooth (capsular) (Avery *et al.*, 1944). Since these early experiments, phase variation in pneumococci has been extensively studied.

The aim of the experiments described in this section was to assess the effect of the capsule on hyaluronate lyase phenotype.

The rationale behind the experiments described here is as follows; where the enzyme is cell associated (cell wall linked via LPXTG motif) and the strain produces a capsule, does the presence of the capsule occlude the enzyme from the surrounding environment? If this were to be the case, variation of capsule production may provide a means of *in vivo* regulation of hyaluronate lyase activity.

3.5.2 – Generation of a capsule negative strain

This experiment required the direct comparison of a capsular strain and its acapsular variant. To achieve this, it was decided to attempt to knock out the capsule locus of a capsular strain of known phenotype.

The serotype three strain 154 was chosen from the phenotypic analysis described in section 3.2. This was a clinical isolate taken from a lower respiratory tract infection. It was proven to be inducibly positive for hyaluronate lyase, with cell associated activity. A serotype three strain was chosen for two reasons; the cellobiuronic acid capsule of this serotype is amongst the more extensive capsules. It was felt that this would be the strongest test for the occlusion hypothesis. Secondly, the phenotype of the serotype three strains is highly characteristic, forming large mucoid colonies on BHI-blood agar, thus facilitating the identification of capsule deficient mutants (looking for small colonies against a

background of large, shiny colonies). First attempts at generating a capsule negative mutant were based on the allelic exchange mutagenesis rationale adopted in generating a Hyl$^-$ mutant.

The genes directing the synthesis of the serotype three capsular polysaccharide (and the other capsular loci studied in molecular detail) lie between two functionally unrelated genes, *dexB* and *aliA*.

Previous work in our lab (Victoria Barcus – personal communication) was performed using a strain kindly donated by Francesco Iannelli of Gianni Pozzi's lab (PR201). In this strain the region between *dexB* and *aliA* had been replaced with a kanamycin resistance cassette. This had previously been used to transform the laboratory strain R6 to kanamycin resistance.

DNA from this derivative strain (R3) was used to amplify the region containing the kanamycin resistance cassette using *dexB* and *aliA* specific primers. This PCR product was approximately 3.5kb in size, and was used in transformation experiments using strain 154.

Despite repeated rounds of transformation using csp1, csp2 and growth phase dependent transformation protocols, no kanamycin resistant acapsular mutants were obtained. It has been observed by other groups that capsular pneumococci are less amenable to transformation by exogenous DNA than acapsular strains (Ravin, 1959, Weiser and Kapoor, 1999), although Watson and Musher were able to select transposon insertion capsular mutants after filter matings with *E. faecalis* (Watson and Musher, 1990). When pneumococci grow as a biofilm, capsular strains have a tendency to become acapsular (J.K. Struthers – Personal communication). Loss of the capsule mediates pneumococcal adhesion to host epithelial cell receptors (see chapter 1, section 1.5.1).

Current work in our lab is addressing the question of variation in capsule production during biofilm growth (R.D. Waite – unpublished data). This investigation employs the Sorbarod biofilm growth model described in chapter 2, section 2.2.5.

Strain ID 154 was passaged through the Sorbarod model, and a number of acapsular variants were obtained from the effluent, some of which reverted to the

capsular phenotype after repeated passage on solid media, plus at least one stable acapsular variant. This stable variant was subsequently renamed AC154.

3.5.3 – Phenotypic characterisation of acapsular variant AC154

The phenotype of the acapsular strain described above was examined using the hyaluronate lyase assays (described in the preceding sections) under non-inducing and inducing conditions. Two induction methods were tested; growing the strain on BHI-blood agar overlaid with 100µl hyaluronic acid [10 mg ml^{-1}] or by inducing for 3 hours in liquid media (C-medium plus hyaluronic acid to a final concentration of [2 mg ml^{-1}]). Figure 3.5.1 below shows the results from this experiment.

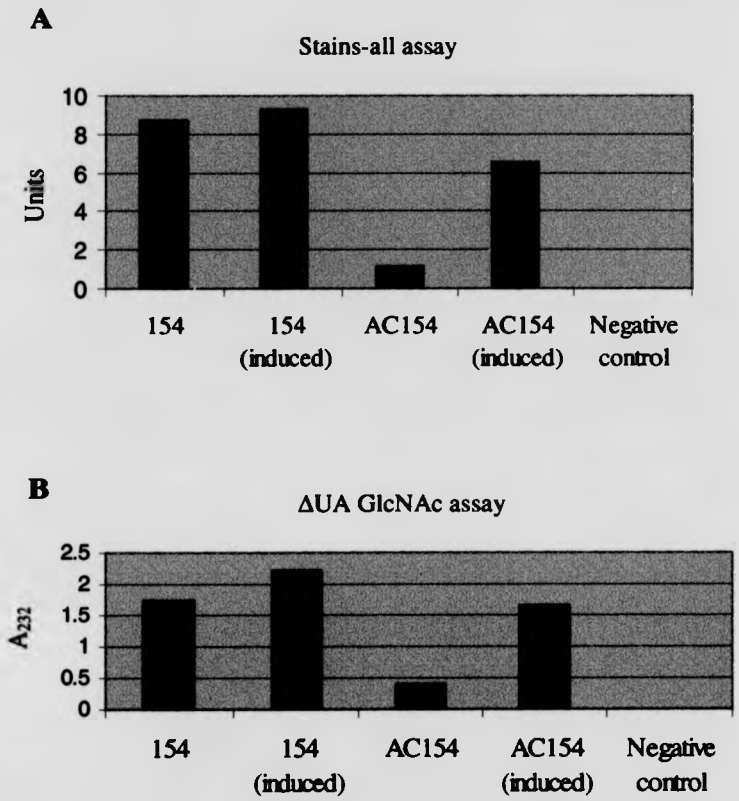


Figure 3.5.1 – Hyaluronate lyase assays of strain 154 and the acapsular derivative, AC154. A; Stains-all assay, B; ΔUA-GlcNAc assay

Both assays show that the activity of the parental (capsular) strain is higher than that of the acapsular derivative, and that induction of hyaluronate lyase activity can be seen with both the parental strain and its acapsular derivative.

3.5.4 – Discussion

The hypothesis described in section 3.5.1 has been disproved on the basis of the results shown in section 3.5.3. Rather than the encapsulated strain having lower activity, it is the acapsular strain that produces less hyaluronate lyase. The fact that

the activity of AC154 is noticeably lower than that of 154 (rather than being the same) is in itself an interesting observation. This will be returned to in section 4.4.5.

The precise nature of the mutation that brought about the stable acapsular phenotype of strain AC154 is not known, so it is not possible to conclude that the hyaluronate lyase phenotype observed is a direct result of the loss of the capsule alone.

3.6 – Characterisation of elements upstream from the published hyaluronate lyase sequence

3.6.1 – Introduction

Analysis of the *hyl* sequence described by Berry *et al.* reveals that all of the codons 5' to their suggested start codon also potentially encode amino acids (Berry *et al.*, 1994). Furthermore, there is no satisfactory ribosome-binding site preceding the suggested start codon. One of the closest known homologues to the pneumococcal enzyme is that found in *Streptococcus agalactiae*. Alignment of their amino acid sequences reveals an additional 164 amino acids at the N terminus of the *S. agalactiae* protein, which includes a cleavable N-terminal signal sequence (Lin *et al.*, 1997)(Gase *et al.*, 1998). Taking these observations together, it seems likely that the available pneumococcal gene sequence is incomplete, and the true 5′ end of the gene and its promoter elements are yet to be described. In their paper Berry *et al.* note the questionable nature of the ascribed promoter in the sequence deposited under accession number L20670.

The available pneumococcal genome databank (http://www.wit.mcs.anl.gov) is not of use in this situation. The hyaluronate lyase ORF lies at the 5′ extremity of one of the unassembled contigs (contig 4167) although an additional 54 bp 5′ to the L20670 start point can be obtained from the contig sequence.

This chapter describes the attempts made to isolate and analyse this region from the laboratory strain R6.

3.6.2 – Characterisation of the region 5′ to the known hyaluronate lyase gene by restriction analysis

Chapter 3.3 describes how the CAT cassette was introduced into the *hyl* ORF. The resultant pneumococcal transformant (P41) now contained a unique marker in the general region of the desired upstream region. The region 5′ to the known

hyaluronate lyase sequence was characterised at low resolution using a combination of restriction digestion and Southern analysis.

The strategy chosen was to use *Bsr*BI, which cuts 119 bp 3' to the end of the CAT insertion (position 1971 with respect to the L20670 sequence). This was used either on its own, or with a range of other enzymes known not to cleave either in CAT or the sequence 5' to the *Bsr*BI site. These enzymes were *Bst*1107I, *Hinc*II, *Hpa*I, *Pvu*II, *Sma*I, *Sna*BI, and *Swa*I. Genomic DNA from strain P41 was digested with these enzymes and separated through 1% agarose. The digest products were transferred to nitrocellulose and probed with Dig labelled CAT PCR product. Chemiluminescent detection of the resultant blot was performed the results of which are shown in figure 3.6.1.

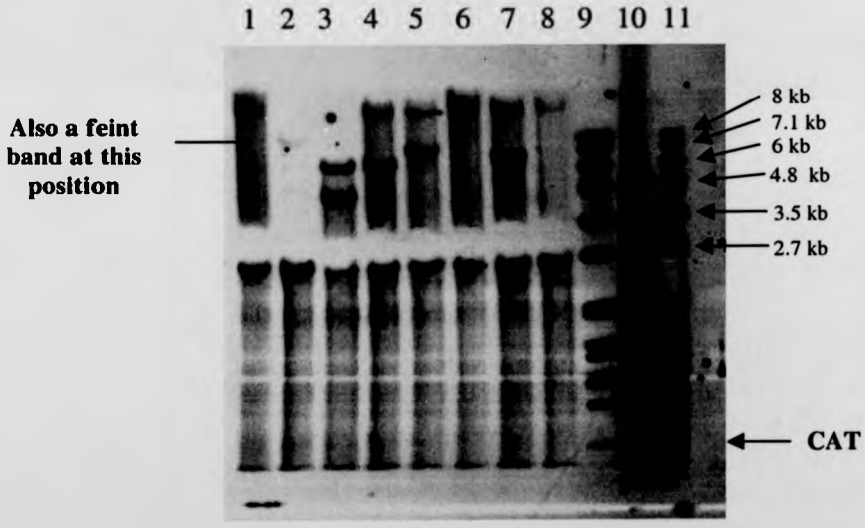


Figure 3.6.1 – Southern blot of P41 digests probed with Dig labeled CAT PCR product. Lanes; 1 = *Bsr*BI only, 2 - *Bsr*BI/*Bst*1107I, 3 - *Bsr*BI/*Hinc*II, 4 - *Bsr*BI/*Hpa*I, 5 - *Bsr*BI/*Pvu*II, 6 - *Bsr*BI/*Sma*I, 7 - *Bsr*BI/*Sna*BI, 8 = - *Bsr*BI/*Swa*I, 9+11 - Dig Markers, 10 - Dig labeled CAT PCR product positive control.

The banding pattern shown in figure 3.6.1 was interpreted thus;

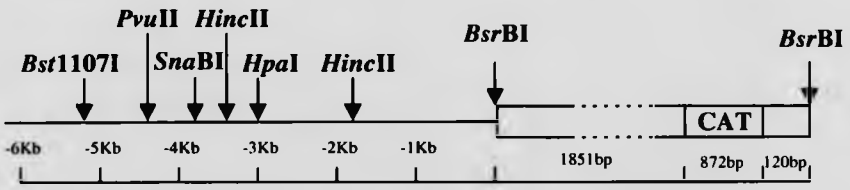


Figure 3.6.2 – Diagrammatic representation of the approximate positions of restriction sites 5′ to the hyaluronate lyase start codon proposed by Berry *et al.* Note that the positions indicated are approximate, within the degrees of accuracy afforded by agarose electrophoresis.

The Southern blot shows that there must be a second *Bsr*BI site close to the known hyaluronate lyase start site since a CAT hybridizing band is seen in all digests. The presence of other bands of higher molecular weight indicate that those products were the result of partial digestion, with cleavage occurring at the *Bsr*BI site 3′ to the CAT insertion site and at one (or in the case of *Hinc*II, two) positions 5′ to the known start point.

3.6.3 – Attempts to clone the hyaluronate lyase promoter from P41

This approach takes advantage of the fact that a selectable marker has already been placed near the region of interest. The digests used for characterisation of the upstream region by Southern analysis were repeated. The resultant restriction products were then tagged with dATP, and cloned into a pre-prepared T-vector (both commercially available vectors and pUC18 prepared in this study were used). Ligation products were transformed into competent *E. coli* (JM109) and cultured on the appropriate selective media.

Despite repeated rounds of the steps described above, no chloramphenicol resistant transformants were obtained. A possible explanation for this is that high levels of transcription from a strong promoter may prove lethal to the *E. coli* host strain.

3.6.4 – iPCR using the pneumococcal transformant P41

As described above, the CAT marker in the R6 derivative P41 proved to be useful in mapping restriction sites 5′ to the known *hyl* start point. It was decided that an inverse PCR (iPCR) approach, using the HIP5.out primer and CAT.rev primer (Cat having gone into the *hyl* gene in the reverse orientation).

Enzymes used in the experiment described above were selected on the basis of the predicted cleavage sites from the data available from the Southern analysis described above. Since *Bsr*BI was now known to cut in the region close to the known start point of the hyaluronate lyase gene, it was decided that this religated product would be easiest to obtain and its approximate size could be verified easily. For completeness, the other restriction enzyme combinations were also included in this experiment. The iPCR rationale is outlined in chapter 2, section 2.2.1.

The result of this experiment is shown in figure 3.6.3.

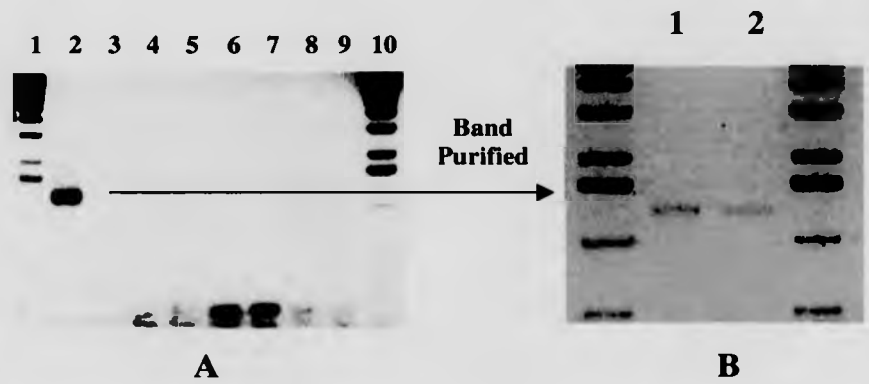


Figure 3.6.3 – A; iPCR using P41 and the digests/primers outlined in the text above. Lanes; 1 - kb Ladder, 2 - *Bsr*BI only, 3 - *Bsr*BI/*Bst*1107I, 4 - *Bsr*BI/*Hinc*II, 5 - *Bsr*BI/*Hpa*I, 6 - *Bsr*BI/*Pvu*II, 7 - *Bsr*BI/*Sma*I, 8 - *Bsr*BI/*Sna*BI, 9 - *Bsr*BI/*Swa*I, 10 - kb ladder. B; Purified band from lane 2 of part A. Lane 1 - Wizard® PCR Preps DNA Purification System, lane 2 - Qiagen QIAquick™ Gel Extraction Kit. 1/20th final recovery volume on gel.

The figure above shows that a single, strong iPCR product was obtained from the *Bsr*BI digest reaction. This band migrates between 1 kb and 1632 bp on the gel (at approximately 1.3-1.4 kb). This was purified using two systems, as shown in part B. This was then used in direct sequencing using the HIP5.out primer. The sequence obtained from this reaction is shown in figure 3.6.4.

5'- GAGCGGAAGAAACGACTACGAATACCATTCAGCAAAGCCAGAAGGA

AGTTCAGTATCAGCAAAGGGATACAAAAAATTTAGTTGAAAATGGTGATT

TTGGTCAGACGGAGGACGGAAGCAGTCCGT**GGACAGGAAGCAAAGCTCAG**

GGGTGGTCAGCTTGGGTAGACCAGAAGAATAGTTCTGCAGATGCCTCAAC

TTGAGTCATTGAGGATCTAAGGCATGGGGACTATCACTATTCTTCTCAAG

CCCTGAGAAATTAAGGGCAGCGGTTCACCGTATGGTTCCCTAA -3'

Figure 3.6.4 – Sequence derived from the iPCR product obtained from the *Bsr*BI digested and ligated P41 genomic preparation described in the text above. Bases indicated in red represent previously unreported sequence. Black and grey bases represent those available from contig 4167 and L20670 respectively. The *Bsr*BI site is underlined. Sequence 5′ to this site returns to the L20670 sequence at position 1971 (not shown).

Figure 3.6.4 indicates that an additional 126 bp of sequence has been obtained from this experiment. Although the total length of additional sequence is quite small (in agreement with the prediction made on the basis of the Southern analysis described above) these findings indicate that the experimental rationale adopted was a valid one.

Analysis of the extra sequence revealed indicate that there would be a stop codon 25 codons upstream from the transcriptional start point predicted by Berry *et al.* This is clearly erroneous, as the published sequence reveals 29 additional codons that encode amino acids. This discrepancy was put down to sequencing errors in the experiment described here.

3.6.5 – Further iPCR experiments using P41

From the data presented above, it can be seen that the iPCR approach was successful in elucidating a further 126 bp upstream from the start of contig 4167 (WIT database sequence). It was decided that this approach should be extended by selecting a different set of restriction digestion conditions and primer sets for the

iPCR reaction. As has already been mentioned, the *Bsr*BI site lies relatively close to the known start-point of contig 4167. As such, this enzyme no longer proves useful in establishing a restriction digest protocol.

The next approach taken was to attempt to anchor the digestion point in the 5′ direction. Once again, it proved possible to draw on the data generated in chapter 3.6.2. The next predicted restriction point in the 5′ direction was *Hinc*II. A *Hinc*II single digest was deemed to be unsuitable, as analysis of the sequence of contig 4167 indicated that there was not another *Hinc*II site for 6.2 kb. This would make any iPCR products obtained with the primers available unfavorably large.

In an attempt to resolve this problem, the 3′ restriction point was defined by selecting known restriction sites of other available blunt cutting enzymes, which lies close to available primers used in the sequencing experiments described in Chapter 3.1.4. This approach is summarised in figure 3.6.5.

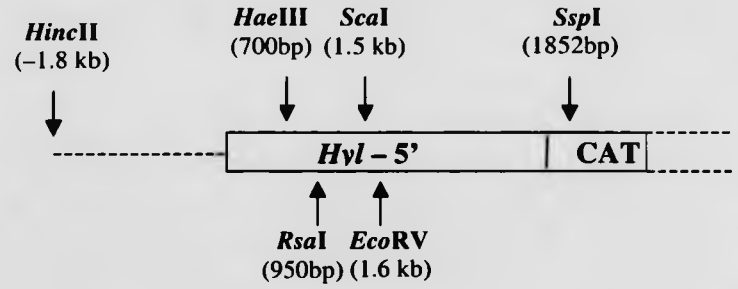


Figure 3.6.5 – Location of restriction sites used in the second round of iPCR described above.

Figure 3.6.5 shows the range of enzymes used in this round of iPCR, and their approximate positions with respect to the first nucleotide of the sequence described by Berry *et al.* (L20670).

An obvious drawback of this approach is that the enzymes cutting 3′ to the start of the gene were not used in the experiment described in section 3.6.2 and as such their ability to cleave within the upstream sequence is unknown. It was decided to proceed with this in any case, as if cleavage were to occur between the proposed *Hinc*II site and the known upstream sequence, this would become apparent from sequence data obtained.

The experiment was carried out as described above, substituting the restriction enzymes appropriately. After ligation the iPCR reaction was performed using the HIP5'.out primer with one of three other enzymes; hyl.for, HSP.F1 and HSP.F1b (see Chapter 2, section 2.1.7, table 2.7 for annealing positions). Of the three primer combinations tried, only the HIP5'.out/HSP.F1 combination gave rise to amplification products. The results of this experiment are shown in figure 3.6.6.

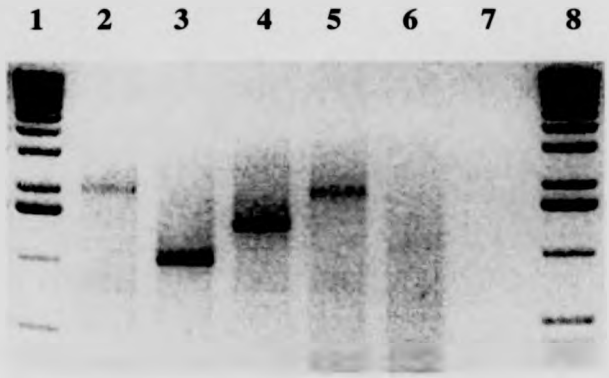


Figure 3.6.6 – iPCR using anchored HincII digest and various other blunt cutting restriction enzymes. Primer pair used was HIP5.out/HSP.F1. Lane 1 - kb ladder, lane 2 - *Hinc*II/*Eco*RV, lane 3 - *Hinc*II/*Hae*III, lane 4 - *Hinc*II/*Rsa*I, lane 5 - *Hinc*II/*Sca*I, lane 6 - *Hinc*II/*Ssp*I, lane 7 - No DNA control, lane 8 - kb ladder.

As can be seen from the figure above, iPCR products were obtained from all reactions except the *Hinc*II/*Ssp*I reaction. These products were purified after agarose gel electrophoresis using the QIAgen QIAquick™ Gel Extraction Kit, and sequencing reactions were performed using the ups.5 sequencing primer, which had been designed against the additional sequence derived previously. Only the *Hinc*II/*Hae*III iPCR product gave rise to readable sequence, which is given in figure 3.6.7, below.

5' – GGCCTCTTTTCTTTTTCTACATTGTAACAGTTTTTACTTTTTTCGTCGAGTC

GGATGGAAGATATTGTTGAAGAATAAGCCCTCAATATGATAGAAACCGTTTACAA

ATAACTTTAAAAATGGTATGATTTAATTAAACCGGTTTACTTTGTTGCTATAAAT

CTCTAATTTGAGATTTAGTTCCTATTGATTTTACAATATGTTTATTGGAGTGTAT

ACATGCAAACAAAAACAAAGAAGCTCATTGTGAGTTTGTCTTCACTTGTTTTATC

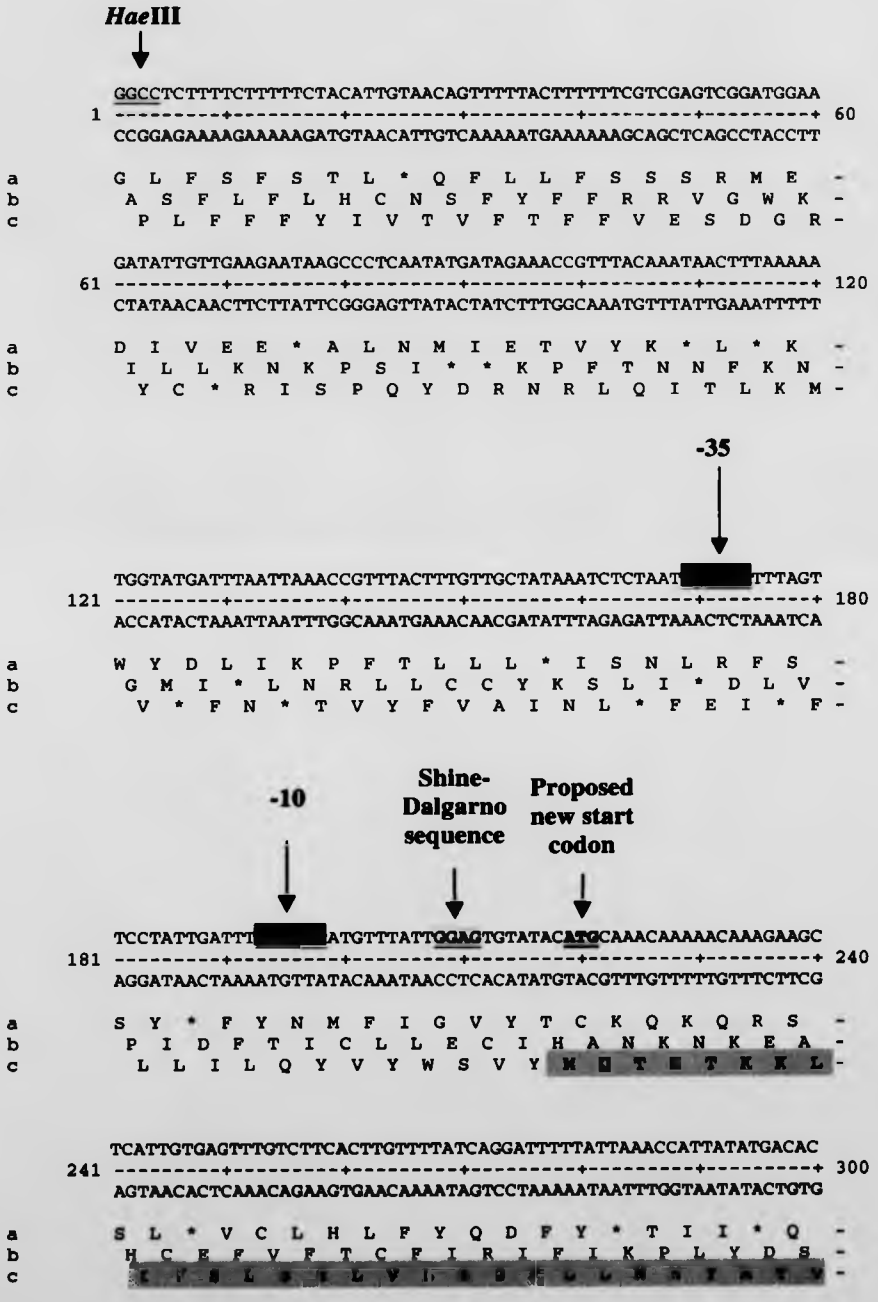
AGGATTTTTATTAAACCATTATATGACAGTTGGAGCGGAAGAAACGACTACGAAT

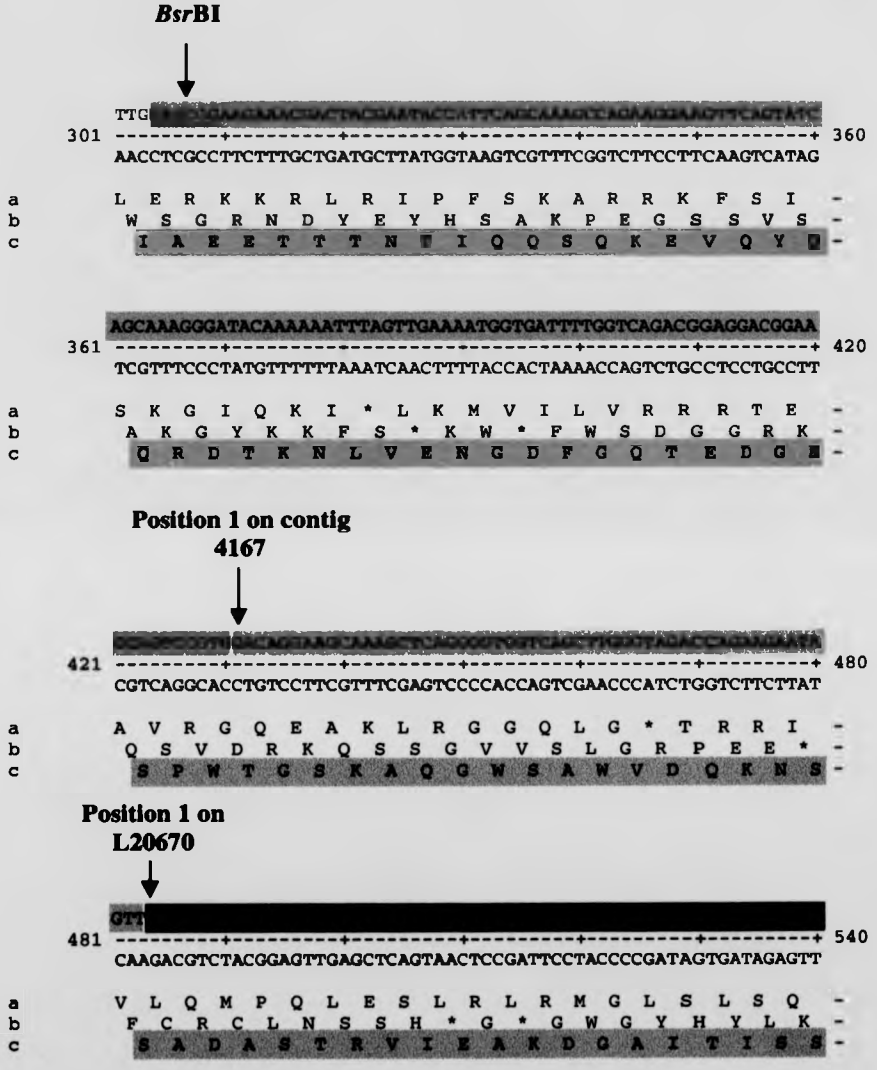
ACCATTCAGCAAGCCAGAAGGAAGTTCAGTATCAGCAAAGGGATACAAAAA - 3'

Figure 3.6.7 – Sequence derived from the iPCR product obtained from HincII/HaeIII digested and ligated P41 genomic DNA. The sequence data obtained in the previous iPCR experiment are shown in red, with the *Bsr*BI site underlined. The additional data obtained in this experiment are indicted in blue, with the single *Hae*III site underlined. Sequence 5' to this point reads back into L20670 at position 682.

As can be seen, this experiment has revealed a further 304 bp of unknown sequence. The known hyaluronate lyase sequence is regained at a *Hae*III site approximately 380 bp into the sequence. Thus, there is a *Hae*III site 431 bp 5' to the first codon of contig 4167, indicating that the attempt to anchor the restriction point at 1.8 kb 5' to the known sequence was unsuccessful due to the presence of this *Hae*III site. Nevertheless, this was still a substantial amount of new sequence data. Sequencing of the other products shown in figure 3.6.6 proved to be unsuccessful.

Figure 3.6.8 places this additional sequence in the context of the known sequence, preliminary upstream sequence data, and the point at which the religated iPCR product runs back into the known hyaluronate lyase sequence. The translation products of the three open reading frames are shown, and the proposed likely ORF and control elements are also given.


HaeIII
-35
-10
Shine-Dalgarno sequence
Proposed new start codon


*Bsr*BI
301
360
AACCTCGCCTTCTTTGCTGATGCTTATGGTAAGTCGTTTCGGTCTTCCTTCAAGTCATAG
361
420
TCGTTTCCCTATGTTTTTTAAATCAACTTTTACCACTAAAACCAGTCTGCCTCCTGCCTT
Position 1 on contig
4167
421
480
CGTCAGGCACCTGTCCTTCGTTTCGAGTCCCCACCAGTCGAACCCATCTGGTCTTCTTAT
Position 1 on
L20670
481
540
CAAGACGTCTACGGAGTTGAGCTCAGTAACTCCGATTCCTACCCCGATAGTGATAGAGTT

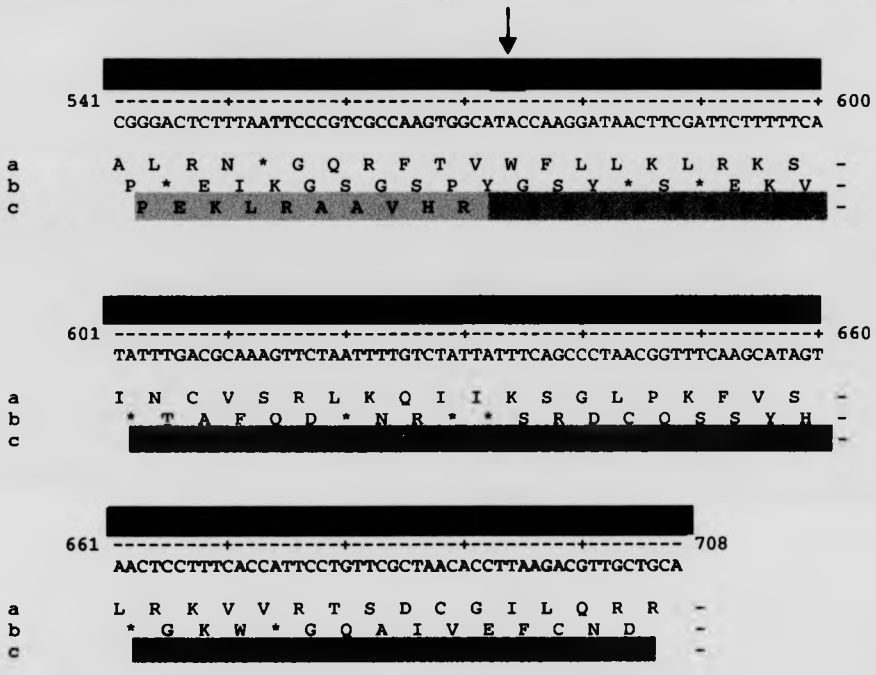


Figure 3.6.8 – Upstream data obtained from the two iPCR experiments compiled and integrated with the known hyaluronate lyase sequence data from the sequence derived by Berry *et al.* (L20670) and on contig 4167 from the WIT database. Key elements are shown, including the *Bsr*BI and *Hae*III sites which limit the extent of information gained from each iPCR experiment. Colour keys are as follows; ■ Known sequence from L20670, ■ Known sequence from contig 4167, ■ sequence derived from the first round of iPCR, terminating in a *Bsr*BI site, unboxed sequence represents data obtained from the second round of iPCR, terminating in a *Hae*III site. The predicted translation product is depicted in gray boxes; ■ ORF translation product predicted by Berry *et al.*, ■ ORF predicted from extended sequence data. –10, -35 and ribosome binding motifs of the predicted true promoter are also indicated.

3.6.6 – Discussion

The diagram above puts data of Berry *et al.*, sequence from the unfinished pneumococcal genome and the data obtained here into context. There is good reason to believe that the true hyaluronate lyase promoter elements are further upstream than those presented in the sequence of Berry *et al.*. Comparison with the sequence of the *S. agalactiae* gene and the data shown here indicate that the hyaluronate lyase ORF proceeds for a further 131 codons before reaching a stop codon. It is reasoned that the true start codon cannot lie 5' to this stop codon, and that the promoter must lie 3' to this region (at around 180 bp on the sequence shown in the diagram above).

Analysis of the data obtained here reveals that there is an excellent candidate for a promoter on the basis of consensus -35, -10 RNA polymerase binding sequences, and their positions relative to one another. These are indicated as red boxes in the sequence shown in figure 3.6.8. The -35 sequence, TTGAGA correlates closely to the consensus sequence – T_{82} T_{84} G_{78} A_{65} C_{54} A_{45}, and is optimally separated from the -10 (Pribnow) box by 18bp. The –10 sequence (TACAAT) agrees closely with the consensus sequence of T_{80} A_{95} T_{45} A_{60} A_{50} T_{96}.

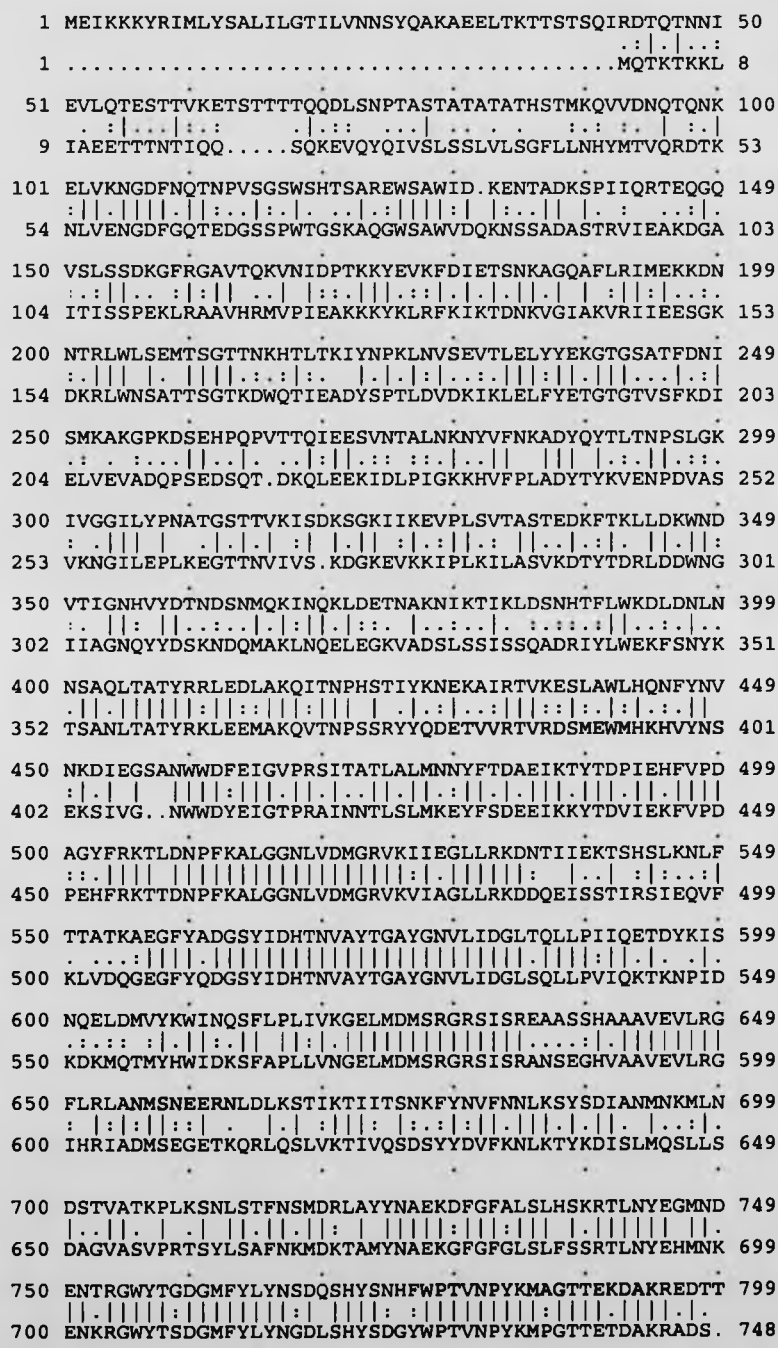
Similarly, there is a good translational consensus at the putative Shine-Dalgarno (or ribosome binding) site, shown in yellow in figure 3.6.8. The consensus sequence of this polypurine stretch is AGGAGG. The sequence presented here is GGAG, representing the core of this sequence. This is separated from the putative start codon by 7 bp.

Stop codons are encountered in all three reading frames in the region of the -35 box, indicating that the start point of the hyaluronate lyase gene cannot lie 5' to the proposed region. The proximity of the promoter elements described above to consensus provides tempting evidence that the promoter truly lies in the proposed region. If the promoter is a very strong one, as implied by the proximity to the consensus elements described, it may provide support for the hypothesis that it has not been possible to clone this element due to lethality caused by overtranscription in the *E. coli* host. Negative control of promoters by trans acting repressor

molecules is common. However, no obvious centres of diad symmetry are evident in the region around the proposed hyaluronate lyase promoter elements.

Translation from the proposed start codon would produce 118 extra amino acids in the translation product. Gase and co-workers found that the *S. agalactiae* protein was 88 residues longer than had previously been described (Gase *et al.*, 1998). The additional residues in the pneumococcal protein would bring the total number of amino acids in each protein into close agreement (1072 residues in the GBS enzyme, compared to 1067 residues in the pneumococcal protein), supporting the hypothesis of an evolutionary relationship between the enzymes.

Alignment of the first two kilobases of the two sequences reveals 51% identity (using the GAP program from the GCG Wisconsin Package), higher than would be expected to occur by chance. Comparison of the amino acid sequences reveals 50% identity and 65% similarity over the length of the whole proteins. Alignment of the N-terminal sequences is less certain, however the N-terminal 31 amino acids share 58% similarity and 35% identity. This is shown in the alignment in figure 3.6.9, where the *S. agalactiae* sequence is the upper sequence, and the pneumococcal sequence is the lower.



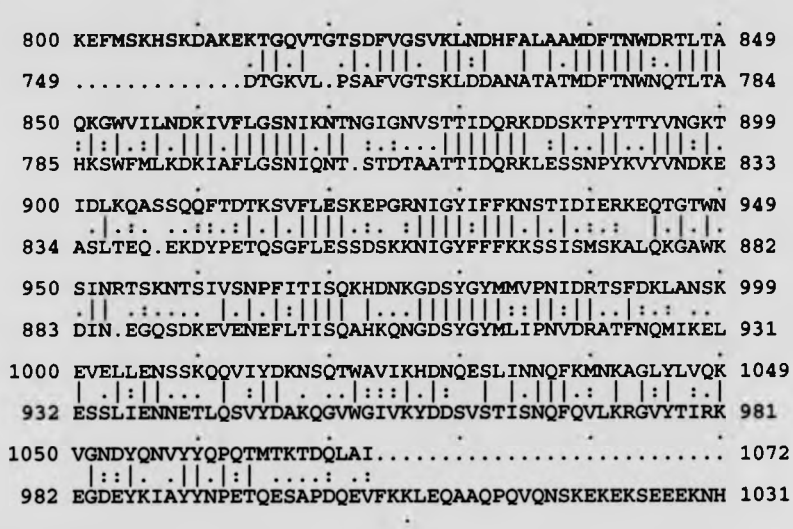


Figure 3.6.9 – Alignment of the extended pneumococcal hyaluronate lyase protein and that of *S. agalactiae* (GBS). The pneumococcal sequence is shown beneath the GBS sequence.

The transcribed protein sequences were analysed by the signal sequence and cleavage prediction methods of McGeoch (McGeoch, 1985), and also the prediction program of Nielsen *et al.* (Nielsen *et al.*, 1997) These facilities can be found at http://psort.nibb.ac.jp/form.html and http://www.cbs.dtu.dk/services/SignalP/#submission respectively. The results of these independent analyses concur, and predict that the additional pneumococcal sequence reported here does indeed contain a cleavable N-terminal signal sequence, with cleavage occurring somewhere between residues 29 and 31. A signal sequence is also found in the GBS amino acid sequence. This is illustrated in figure 3.6.10.

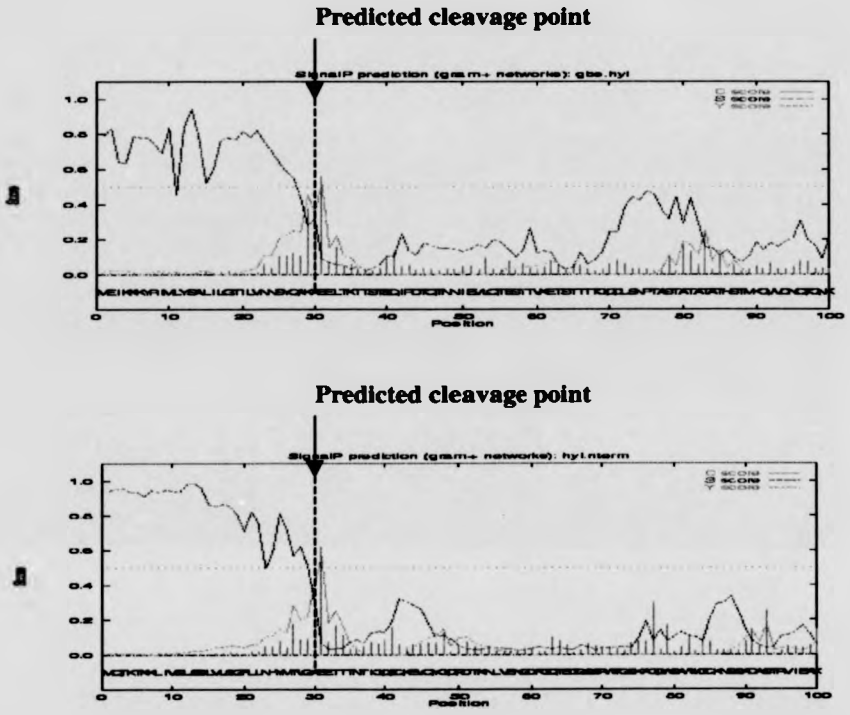


Figure 3.6.10 – Signal sequence prediction by the method of Nielsen *et al.* The desired profile of a signal sequence includes high S score (blue) before the cleavage point, low after and a high C score (green) immediately after the cleavage site. The GBS sequence is shown at the top, with the pneumococcal sequence at the bottom.

Analysis of the ORF ascribed by Berry *et al.* does not lead to the conclusion that the protein contains such a signal sequence. The presence of such a signal sequence in the extended protein is considered to provide a basis for the observation that hyaluronate lyase activity is high in the supernatant fraction of the protein encoded by the *hyl* allele in strain ID 922, with two stop codons occurring prior to the region encoding the cell wall anchor motif.

This is also likely to account for the supernatant activity noted by Berry *et al.* although not by the LytA dependent mechanism, which they propose (Berry *et al.*, 1994).

3.6.7 – Exploration of regions 5′ to the hyaluronate lyase promoter

Genome sequencing efforts by another group have indicated that the pneumococcal *hyl* locus lies 3' to the pneumococcal glutathione peroxidase gene (*gpx*) (unpublished - anonymous communication).

Analysis of the WIT database shows that the *gpx* gene lies at the 3' end of contig 4154, with its ORF on the minus strand with respect to the majority of the genes on that contig. This arrangement is depicted in figure 3.6.11.

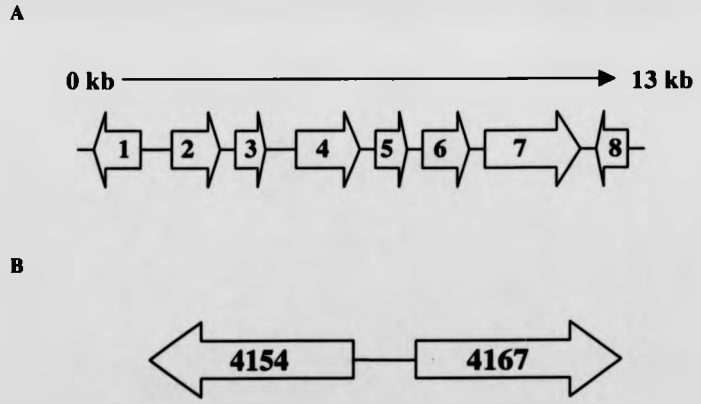


Figure 3.6.11 – Structure of contig 4154 and the proposed localisation of this contig with respect to contig 4167, bearing the *hyl* gene. A: Genes on contig 4154; 1 = Hypothetical cytosolic protein, 2 = 6-phospho-β-glucosidase, 3 = PTS – cellobiose specific IIB component, 4 = putative cel operon regulator, 5 = PTS cellobiose specific IIA component,6 = PTS cellobiose specific IIC component, 7 = α-glucosidase, 8 = glutathione peroxidase. B: Proposed localisation and orientation of contigs 4167 and 4154 with respect to one another.

PCR primers were designed to read out from *gpx* in the 3' to 5' direction with respect to the open reading frame of the gene, and were used in conjunction with *hyl* specific primers, reading outwards with respect to the hyaluronate lyase ORF (note that the ORF's of the two genes mentioned lie on the + and – strands with respect to one another). The experiment was conducted with 11 strains of different hyaluronate lyase phenotypes (see chapter 3.2). The results of this PCR experiment are shown in figure 3.6.12.

Figure 3.6.12 - Testing the hypothesis that *gpx* lies 5' to *hyl* in the opposite orientation. A: The proposed location of the genes, the primers used and information known about the relative distances involved in bp. The unbroken line between the OFR's (indicated by boxes) represents the known upstream sequence, described in section 3.6.5. The dashed line represents the putative, unknown sequence. B: Results of PCR using the gpx.out and ups5'.out primer pair. Lanes; 1 - Kb ladder, 2 – strain ID's 213, 3 - 318, 4 - 39, 5 - 31, 6 - 91, 7 - 87, 8 - 324, 9 - 329, 10 - 315, 11 - 30, 12 - 205, 13 - no DNA control, 14 - Kb ladder.

As can be seen, a product has been amplified from all strains tested, confirming the proposed location of the two genes with respect to one another. Size estimation from the ladder indicates that the product is approximately 700 bp in length. Given the information known about the primer binding sites this would mean that the unknown intervening sequence is approximately 200 bp in length. To confirm this, some of the products were selected for direct sequencing using the ups5'.out and gpx.out primers. The results of this experiment, shown in figure 3.6.13 below,

confirm that contig 4154 does indeed lie 5′ to contig 4167 with respect to the *hyl* ORF.

5′ - CGTGGGATAAACTGATTGCAAGGGAAATCCAATATTTCGAAGCCCTGATCTTGTAACGT
TCATAGAGTTCTTGAAGTCCCTGGTACTGGGGCGTTAAACCACATCCAGTAGCAGTGTT
GACAATCAAGAGAACTTTGCCACGATAGCTATCCAAGGGAGTTGCTTGGTTGTTTTGGT
TCAAAACGGAAAAATCATAAAGTGAAGTCATGAGGGCCTCTTTTCTTTTTCTACATTGT
AACAGTTTTTACTTTTTTCGTCGAGTCGGATGGAAGATATTGTTGAAGAATAAGCCTTC
AATATGATAGAAACCGTTTACAAATAACTTTAAAAATGGTATGATTTAATTAAACCGGT
TTACTTAGTTGCTCTAAATCTCTAATTTGAGATTTAGTTCCTATTGATTTTACAATATG
TTTATTGGAGTGTATACATGCAAACAAAAACAAAGAAGCTCATTGTGAGTTTGTCTTCA
CTTGTTTTATCAGGATTTTTATTAAACCATTATATGACAATTGGAGCGGAAGAAACGAC
TACGAATACCATTCAGCAAAGCCAGAAGGAAGTTCAGTATCAGCAAAGGG - 3′

Figure 3.6.13 – Sequence reading into contig 4154 from upstream elements of hyaluronate lyase. Red sequence indicates the 5′ end of contig 4154. Blue sequence indicates *hyl* upstream sequence data obtained as described in the preceding sections. Key restriction sites used in iPCR experiments are underlined with the proposed *hyl* start codon in bold.

The sequence data above show that we gain an additional 179 bp, in close agreement with the predicted size of the intervening region from the experiment outlined in figure 3.6.12. Analysis of this sequence does not reveal any further promoter like elements with respect to the *hyl* ORF. However, viewed in the opposite orientation, it seems as if the *gpx* gene had suffered the same fate as the *hyl* gene, in that the available sequence data proved to be truncated at the 5′ end. Extension of this sequence yields further codons encoding amino acids, leading to a proposed promoter region. It is thought that this extension of the *gpx* sequence is valid on the basis of similarity of the pneumococcal glutathione peroxidase protein with homologues from *S. pyogenes* and *B. subtilis*. These points are illustrated in figures 3.6.14 and 3.6.15.

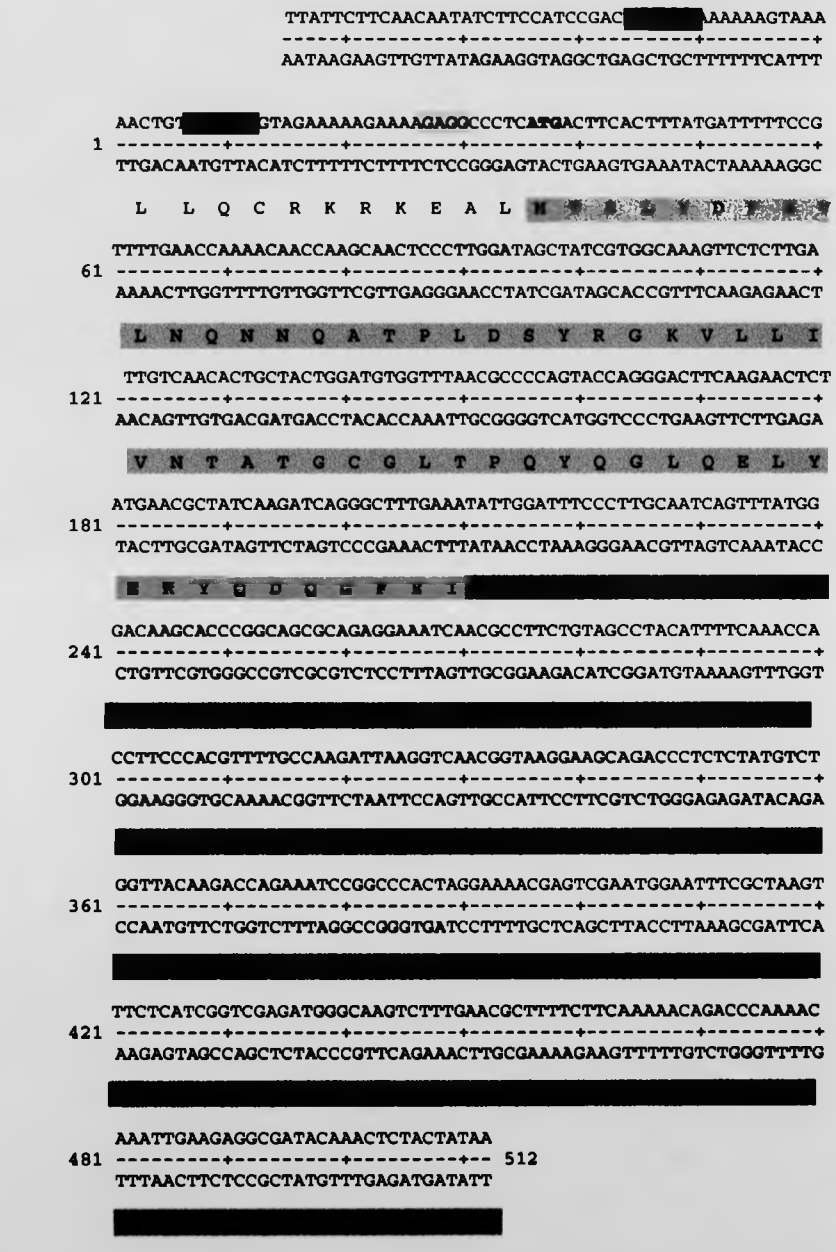

TTATTCTTCAACAATATCTTCCATCCGAC[redacted]AAAAAGTAAA
AATAAGAAGTTGTTATAGAAGGTAGGCTGAGCTGCTTTTTTCATTT
AACTGT[redacted]GTAGAAAAAGAAAAGAGGCCCTCATGACTTCACTTTATGATTTTTCCG
1
TTGACAATGTTACATCTTTTTCTTTTCTCCGGGAGTACTGAAGTGAAATACTAAAAAGGC
L L Q C R K R K E A L M
TTTTGAACCAAAACAACCAAGCAACTCCCTTGGATAGCTATCGTGGCAAAGTTCTCTTGA
61
AAAACTTGGTTTTGTTGGTTCGTTGAGGGAACCTATCGATAGCACCGTTTCAAGAGAACT
L N Q N N Q A T P L D S Y R G K V L L I
TTGTCAACACTGCTACTGGATGTGGTTTAACGCCCCAGTACCAGGGACTTCAAGAACTCT
121
AACAGTTGTGACGATGACCTACACCAAATTGCGGGGTCATGGTCCCTGAAGTTCTTGAGA
V N T A T G C G L T P Q Y Q G L Q E L Y
ATGAACGCTATCAAGATCAGGGCTTTGAAATATTGGATTTCCCTTGCAATCAGTTTATGG
181
TACTTGCGATAGTTCTAGTCCCGAAACTTTATAACCTAAAGGGAACGTTAGTCAAATACC
E R Y Q D Q G F E I[redacted]
GACAAGCACCCGGCAGCGCAGAGGAAATCAACGCCTTCTGTAGCCTACATTTTCAAACCA
241
CTGTTCGTGGGCCGTCGCGTCTCCTTTAGTTGCGGAAGACATCGGATGTAAAAGTTTGGT
[redacted]
CCTTCCCACGTTTTGCCAAGATTAAGGTCAACGGTAAGGAAGCAGACCCTCTCTATGTCT
301
GGAAGGGTGCAAAACGGTTCTAATTCCAGTTGCCATTCCTTCGTCTGGGAGAGATACAGA
[redacted]
GGTTACAAGACCAGAAATCCGGCCCACTAGGAAAACGAGTCGAATGGAATTTCGCTAAGT
361
CCAATGTTCTGGTCTTTAGGCCGGGTGATCCTTTTGCTCAGCTTACCTTAAAGCGATTCA
[redacted]
TTCTCATCGGTCGAGATGGGCAAGTCTTTGAACGCTTTTCTTCAAAAACAGACCCAAAAC
421
AAGAGTAGCCAGCTCTACCCGTTCAGAAACTTGCGAAAAGAAGTTTTTGTCTGGGTTTTG
[redacted]
AAATTGAAGAGGCGATACAAACTCTACTATAA
481
TTTAACTTCTCCGCTATGTTTGAGATGATATT
512
[redacted]

Figure 3.6.14 – Translation of the entire region encompassing the pneumococcal *gpx* ORF. ■ –35 and –10 control regions, □ Shine-Dalgarno sequence, ■ amino acid sequence described on contig 4154, ■ additional amino acids obtained from sequence data presented here.

Figure 3.6.15 (below) shows the alignment of this extended pneumococcal Gpx protein with those of other organisms.

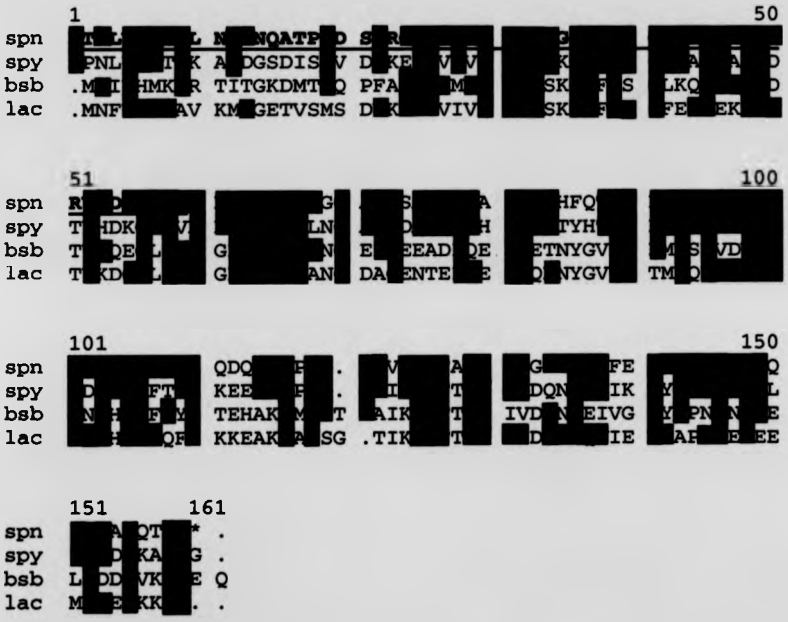


Figure 3.6.15 – Alignment of the pneumococcal glutathioine peroxidase amino acid sequence with those of other bacteria. Only areas of amino acid identity are highlighted. The additional pneumococcal protein amino acids are undelined and in bold typeface. spn = *S. pneumoniae*, spy = *S. pyogenes*, bsb = *B. subtilis*, lac = *L. lactis*.

As can be seen, the additional pneumococcal sequence agrees closely with one or more of the related protein sequences. This strongly suggests that the proposed extension of the *gpx* gene is correct.

In conclusion, it seems that it has been possible to describe the 5' termini and promoter elements of the *hyl* and *gpx* genes, and evidence has shown that the contigs 4154 and 4167 are located next to one another in the pneumococcal genome.

4.1. – Variation at the locus encoding the pneumococcal hyaluronate lyase, *hyl*

Chapter 3, section 3.1 describes the analysis of genetic variation at the *hyl* locus by HRRA. PCR amplification of a 2709 bp fragment of this gene was successful from 115/122 (94.2%) strains tested. Those strains that were PCR negative were screened by Southern analysis, confirming the presence of the gene in these strains.

Berry *et al.* reported that *hyl* is present in the pneumococcal genome as a single copy (Berry *et al.*, 1994). The Southern analysis performed here confirms this observation using a larger number of strains from diverse geographical locations (Spain, UK, Kenya) and different disease states (carried, septicaemia, pneumonia) over a time period ranging from 1988/89 until 1995.

The HRRA studies of *hyl* revealed that divergence at this locus was not notably higher than that of the housekeeping loci *hexA* and *recP*. Low nucleotide divergence was also observed with the pneumococcal *lytA* gene, although this example was complicated by evidence of recombination with loci encoding bacteriophage associated autolysins (Whatmore and Dowson, 1999). Direct sequence analysis of selected *hyl* loci revealed drawbacks associated with the HRRA technique. The prediction of nucleotide divergence estimated by the algorithm of Nei and Li is in some cases inaccurate, however comparisons between the loci studied are thought to be valid since they are likely to undergo variation in the same manner.

So, *hyl*, unlike other pneumococcal loci that are clearly under strong selective pressure (e.g.; the PBP's) is not highly variable. This is an interesting observation; if the protein is cell associated (as the C-terminal cell wall anchor implies), low variation may indicate that the protein is not highly antigenic.

Paton *et al.* have reported that they were unable to demonstrate protection in a murine intra-peritoneal immunisation/challenge model, suggesting low antigenicity, although this was perhaps not the model of choice, since hyaluronate lyase may not be a virulence factor during intra-peritoneal challenge (Paton *et al.*, 1997). Intra-nasal challenge data has not been reported.

4.2 – Direct sequence analysis of selected alleles

Analysis of the translation products of the alleles sequenced show that non-synonymous alterations to the published sequence seem to be preferentially located in the α-domain of the molecule. There are two possible reasons for this, which are not mutually exclusive. The first is that variation of components of the β-domain is not tolerated due to steric restrictions on the topology of this domain. The second is that the α-domain, extending further into the extracellular environment, is subject to any antigenic recognition that may exist.

For a robust interpretation of the observed amino acid changes in terms of antigenic variation, a detailed map of the structure reported by Li *et al.* would be required in order to localise the changes observed to the respective structural elements. One tempting observation has been made on the basis of the structural information available in the Li paper. It is possible to determine the location of one of the more exposed loops (αL2) at the extremity of the α-domain to the region of amino acids 225-235 of the sequence of Berry *et al.* In this region there are 3 different amino acid alterations seen, which affect up to 7/9 of the alleles sequenced. If antigenic selection were driving the limited amount of variation seen, it might be expected to manifest itself in clusters of alterations in surface exposed structures such as this.

Transformation mediated homologous recombination is a well-documented mechanism for genetic variation within *Streptococcus pneumoniae*. Maximum chi-squared analysis of the distribution of nucleotide alterations across the *hyl* alleles studied indicates that although the overall level of divergence between the alleles is low, there is statistically significant evidence for what appears to be recombination between *hyl* alleles. Molecular rearrangements of this nature would account for the higher number of *hyl* alleles found by HRRA compared to the housekeeping loci studied.

This does not exclude the possibility of inter-specific recombination, however a PCR based screen for *hyl* homologues in representatives of atypical pneumococci

and related oral streptococci did not reveal any potential donors for divergent regions.

Taken together these findings indicate that recombination is restricted to the *hyl* alleles found in pneumococci. The inability to find *hyl* homologues in *S. mitis*, *S. oralis* or atypical pneumococci (which have previously been identified as *ply* and *lytA* positive (Whatmore *et al.*, 2000)) is another example of the restriction of this gene to the 'typical' members of this group. Although other streptococci are phenotypically Hyl^+, their *hyl* loci must be sufficiently divergent from the pneumococcal locus to render them negative in PCR and Southern based screens using probes based on the pneumococcal locus (this thesis and S. Widdison – unpublished data). Given the level of genetic promiscuity of the streptococci, it is perhaps surprising to find such restriction of this locus to species type, even to the extent of its absence from atypical isolates bearing other pneumococcal characteristics. Southern analysis is in progress within the group to address this question.

These observations have possible implications for the design of next generation conjugate vaccines. As mentioned in section 1.5.2, it is believed that conjugation of pneumococcal capsular polysaccharides to protein carriers will generate improved, thymus-dependent immune responses in key target populations.

The findings of this study suggest that hyaluronate lyase is a potential candidate for inclusion in such vaccines. It is a protein that appears to undergo limited variation at the amino acid level (notably in the β-domain).

4.3 – Phenotypic analysis of hyaluronate lyase production

The publication of Kostoyukova *et al.* implied that hyaluronate lyase production was greater from pneumococci isolated from the CSF of meningitis patients than from carried isolates (Kostoyukova *et al.*, 1995). The authors did not describe efforts to exclude large numbers of clonally related isolates from their study, or conduct a rigorous statistical interpretation of their results. The investigation was emulated using pneumococci isolated from a wider range of disease states, and which were

varied on the basis of as many factors as possible to avoid the inclusion of multiple clonal isolates.

For statistical robustness, two independent assay systems were chosen on the basis that if one proved to be unreliable, then the other could serve as a backup. The assays approached the activity of hyaluronate lyase by measuring both the degradation of the HA substrate and the accumulation of reaction products.

The results of these assays (conducted on 185 strains) were compared using the model II anova technique, where isolates were grouped by their site of isolation. Pairwise post-hoc tests were performed to identify the source of variance identified in the Anova procedure. These tests showed that in some cases the mean titres of the isolates taken from the lower respiratory tract were lower than those of other groups of isolates. There was no evidence to support the findings of Kostoyukova *et al.* The variance from the means was found to be high within individual groups.

This heterogeneity reflects the findings of Reinholdt and Kilian, who examined the IgA protease phenotypes of a number of pneumococcal isolates and found considerable variability in activity (Reinholdt and Kilian, 1997).

There is the obvious notion that comparison of phenotype between strains studied *in vivo* and *in vitro* is complicated by the acceptance that pathogens regulate phenotype in response to their environment (as with competence induction, for example). It was felt that exposure of pneumococci to the hyaluronate lyase substrate might modulate the production of the enzyme.

In order to address this, all isolates were incubated in defined media containing HA prior to phenotypic analysis. It was found that in the majority of cases (65%) activity was higher after this treatment, that is, pneumococci are apparently inducible for hyaluronate lyase. This effect was shown to be statistically significant by the paired T-test.

The conclusion from this stage of the work described here is that care must be taken when ascribing phenotypic traits *in vitro* to behaviour *in vivo*. These experiments have endeavoured to obtain the 'true' hyaluronate lyase phenotype in the laboratory by induction of hyaluronate lyase by exposure of pneumococci to HA.

4.4 – Hyaluronate lyase mutants

The first *hyl* mutant encountered in this study was strain 619 (ID 53), found in the HRRA study. This strain has a 400bp deletion in the *hyl* gene, and was shown to be phenotypically Hyl^-. This strain was isolated from a Spanish patient suffering vaginal sepsis in 1987. The unusual site of isolation may imply trophic restriction placed upon this strain by its inability to degrade hyaluronic acid.

Since the genetic background of the Hyl^- mutant strain 619 was unknown, a defined Hyl^- strain was constructed by interposon mutagenesis.

This mutant was also unable to degrade HA, thus confirming that there is no functional redundancy with respect to hyaluronate degradation by the pneumococcus.

No other mutants were found which were pnenotypically Hyl^-, although 2 mutants that were Hyl^+ were identified during the sequence analysis.

The first was from a Kenyan throat isolate (strain 86012 - ID 307), harbouring *hyl* allele 22. This allele contains a very unusual pair of inserted sequences that together restore the reading frame of the gene (see chapter 3, section 3.1.6). This strain was shown to be phenotypically Hyl^+.

Either insertion alone would result in the introduction of variant amino acids, followed by a stop codon at a position N-terminal to that expected in the wild type.

Analysis of allele 2 also reveals a truncation N-terminal to either of those introduced by either of the insertion events found in allele 22.

This mutant was identified from the carried UK strain 12B-13E (ID 922). This allele contains two stop codons 3' to the region encoding the cell wall anchor motif. Translation from this ORF results in the production of an enzyme lacking the C-terminal region containing the LPXTG cell wall anchor motif.

This has interesting implications for modulating the activity of the enzyme, especially in the light that the protein possesses an N-terminal signal sequence (see chapter 3, section 3.6 and this chapter, section 4.7).

Hyaluronate lyase assays using this strain showed negligible activity associated with the cell-associated fraction, but high supernatant activity, indicating that the enzyme is indeed exported directly into the extracellular environment.

This representative of allele 2 (the most common allele found) was identified from a successful serotype 3 clone (designated clone C) described by Muller-Graf *et al.* (Muller-Graf *et al.*, 1999). This allele was compared with other strains (of the same HRRA *hyl* genotype) identified as bearing from within the same clonal group, and strains which are unrelated to this clone. It was found that the truncation was restricted to the clonally related strains, with the clonally unrelated strains bearing the wild-type sequence at this point.

The interpretation of these findings is that the mutation has occurred at some point in the evolutionary history of the progenitor strain of clone C, which has been amplified with the clonal expansion of this group of isolates. Clone C is a large national clone whose expansion indicates that the loss of the hyaluronate lyase cell wall anchor is not detrimental to the success of the clone.

Returning to the insertions found in allele 22. It is not possible to say whether the insertions were introduced to this allele during the same, or independent mutation events. It is interesting to speculate that there has been a selective pressure to restore the cell-wall associated phenotype, hence the second duplication which restores the open reading frame required for the translation of a functional cell-wall anchor. Such restoration of phenotype by a duplication event has also been seen in the locus encoding photosystem II of *Synechocystis,* under nutrient limiting conditions (Kless and Vermass, 1995). Since allele 2 shows that hyaluronate lyase activity is not abolished by N-terminal truncation, we must assume that maintenance of hyaluronate lyase at the cell surface is a conditional, rather than absolute requirement.

4.5 – Pneumococcal growth on hyaluronate

The nature of the host/pathogen relationship is, by necessity, exquisitely complex. While hyaluronate lyase activity is considered to be involved in penetration

of the host tissues during invasive disease, there is also evidence that the enzyme contributes to pneumococcal persistence by utilising host hyaluronic acid as a carbon source (Chapter 3, section 3.4).

Various Gram negative organisms, including *Staphylococcus epidermidis, Streptococcus pyogenes* and *S. intermedius* have the ability to utilise hyaluronic acid as a sole carbon source (Costagliola *et al.*, 1996, Homer *et al.*, 1997). The approach described in chapter 3, section 3.4 was successful in showing that HA was able to support growth as a sole carbon source with various strains of *S. pneumoniae*. Experiments with the isogenic *hyl* mutant and its parental strain indicated that this mode of growth was dependent on hyaluronate lyase.

In the case of *S. intermedius*, this mode of growth is catabolite repressed by glucose. This was not thought to operate in the pneumococcal system, as induction of hyaluronate lyase was unaffected using induction media containing glucose (data not shown), however it cannot be ruled out that other enzymes in the HA utilisation pathway are affected by the presence of glucose.

The reaction product of hyaluronate lyase is an unsaturated disaccharide that is likely to accumulate in the extracellular environment, given the proposed location of the enzyme on the cell surface. Events downstream of hyaluronate lyase function are proposed to include an exoglycosidase activity and a permease activity, the order of these is unknown in the pneumococcal system, but is speculated upon below.

Studies using other metabolites showed that exogenous GlcNAc, but not glucuronic acid supported growth with these strains. Growth on N-acetylglucosamine (GlcNAc) is proposed to proceed by the interconversions outlined in figure 4.1. Note; The following references to the pneumococcal genome sequence were made using the genome resource at http://www.wit.mcs.anl.gov.

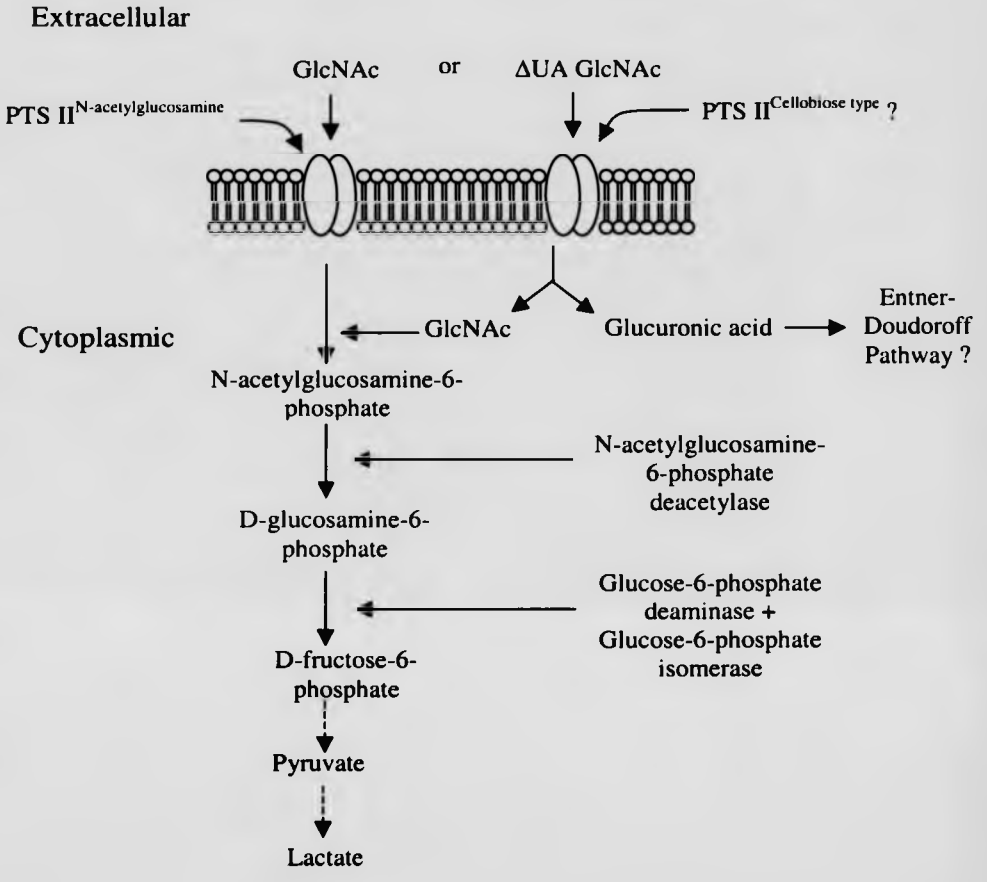


Figure 4.1 – N-acetylglucosamine catabolic pathway. Pneumococci are homofermentative, able to degrade glucose via the fructose intermediate. For brevity, only the steps leading to fructose-6-phosphate are shown in detail. Reactions downstream from this point are core metabolic activities, indicated by dashed lines. See text for details regarding uptake and proposed metabolic pathways.

The conversion of N-acetylglucosamine to fructose-6-phosphate involves internalisation, phosphorylation, deamination and deacetylation before an isomeric conversion step. This metabolic route was once perceived to be an exclusively Gram-negative trait, however similar conversions are known to occur in Gram-positive organisms, see below. In *E.coli* the interconversions outlined in figure 4.1 are carried

out by enzymes encoded by the *nag* operon. The first two steps are carried out by a multipass transmembrane protein of the sugar phosphotransferase class, encoded by *nagE* (Peri and Watgood, 1988). This enzyme was first described in *E.coli* and numerous homologues exist in *S. pneumoniae* (see http://www.wit.mcs.anl.gov)

A pneumococcal homologue of the glucose-6-phosphate isomerase (encoded by *nagC* in *E.coli*) has also been found.

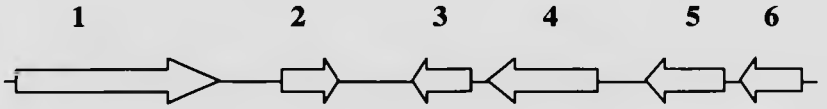
While a homologue for the deacetylase encoding gene has not been found in *S. pneumoniae*. The available pneumococcal genome sequence is incomplete at the time of writing, so it is possible that the gene exists in a region of the pneumococcal chromosome that is not yet available in the public domain. Given the pneumococcal growth phenotype, this seems plausible.

Homer *et al.*, have shown that both *Streptococcus mutans* and *S. oralis* utilise GlcNAc in the manner outlined above (Homer *et al.*, 1993, Homer *et al.*, 1996), leading to the suggestion that pneumococci share this metabolic trait.

A different hypothesis is that the conversion of GlcNAc to pyruvate proceeds via an alternative pathway, with glucosaminate as an intermediate (Michal, 1993).

This is not thought to be the case for two reasons; there is no homologue of the key enzyme, aminodeoxygluconate dehydratase in the available pneumococcal genome databank. Also, the production of hydrogen peroxide during the conversion of D-glucosamine to glucosaminate is unfavourable since pneumococci are catalase negative organisms.

Interestingly, two key enzymes in the glucosaminate pathway are found in the pneumococcal genome on the same contig (4167) as the hyaluronate lyase gene, namely 2-keto-3-deoxygluconokinase and 2-keto-3-deoxygluconate aldolase. The organisation of genes on contig 4167 is illustrated in figure 4.2.



0kb 7kb

ORF assignments:
1 – Hyaluronate lyase
2 - Transposase (putative assignment)
3 - 2-keto-3-deoxyphosphogluconate aldolase
4 - 2-keto-3-deoxy-D-glucokinase
5 – Galactose-6-phosphate isomerase (LACB subunit)
6 – Gluconate-5-dehydrogenase

Figure 4.2 – ORF organisation of contig 4167.

The enzymes encoded by ORF's 3 and 4 are required for uronic acid catabolism via the Entner-Doudoroff pathway. This is suggestive of a role in the utilisation of glucuronic acid.

Although the addition of exogenous GlcUA is unable to support growth, if the cell is able to transport the disaccharide product of hyaluronate lyase, these intracellular enzymes may enable the utilisation of both saccharide moieties of HA. The Entner-Doudoroff pathway is depicted in figure 4.3.

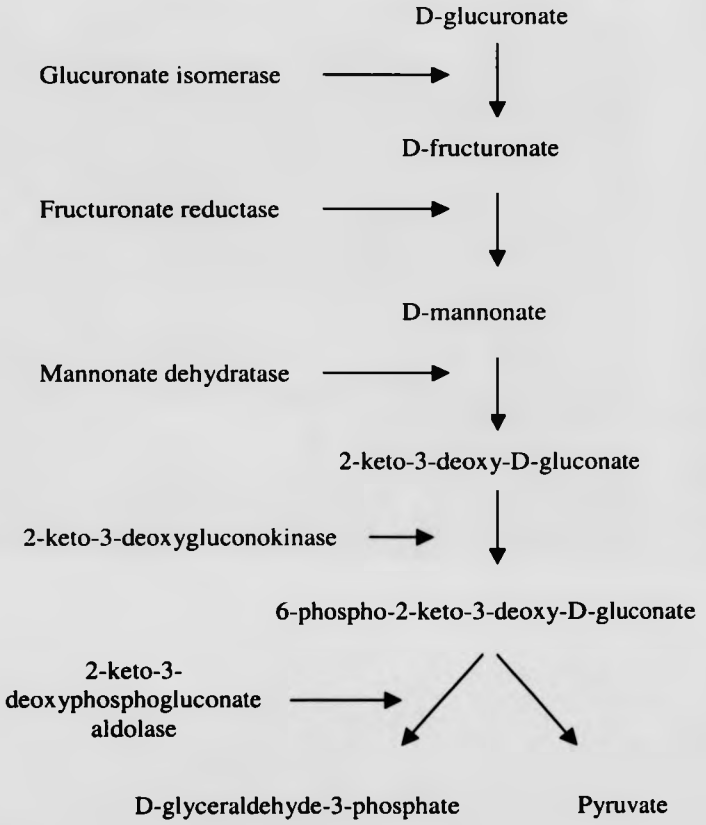


Figure 4.3 – Conversion of glucuronic acid to glyceraldehyde-3-phosphate and pyrvate via the Entner-Doudoroff pathway. Genes required for growth on glucuronic acid via the Entner-Doudoroff pathway have been reported in various organisms including *E. coli, B. subtilis, H. influenzae* and *P. aeruginosa.*

Of the five enzymatic stages in the pathway outlined in figure 4.3, two are notably absent from the currently available genome sequence data. These are gluconate isomerase and mannonate dehydratase. However, the pneumococcal genome does contain a homologue to the *E. coli* fructuronate reductase.

The presence of the N-acetylglucosamine transmembrane transporter protein indicates that the pneumococcus possesses the necessary apparatus to take up this metabolite under the appropriate conditions. This may or may not be related to its mode of growth on hyaluronate however (for example, it may be required for pneumococcal survival in the bloodstream during bacteraemia). Cleavage of the hyaluronate lyase product during HA utilisation by *S. intermedius* is carried out by an intracellular enzyme (Homer *et al.*, 1997), requiring the uptake of the disaccharide by the cell.

The results from Chapter 3, section 3.6 reveal that contig 4167, which contains *hyl* lies downstream of contig 4154. This contig contains loci encoding a cellobiose spceicfic PTS transporter (see chapter 3, section 3.6.7, figure 3.6.11).

The repeat unit of cellobiose is very similar to that of hyaluronic acid, and it is tempting to speculate that this transporter may actually be involved in the internalisation of the pneumococcal hyaluronate lyase reaction product.

The observation that HA can support growth of some pneumococcal strains provides compelling evidence to suggest that the status of this virulence factor should be broadened to encompass its contribution to persistence in the host, as well as its role as a spreading factor.

4.6 – Effect of capsule on Hyl phenotype

The work described in chapter 3, section 3.5 was conducted in order to assess the effect of capsule loss on hyaluronate lyase phenotype. The starting hypothesis was that the presence of large amounts of capsular polysaccharide (as with the serotype 3 capsule) might modulate the activity of cell associated hyaluronate lyase by occluding the enzyme from its substrate.

In fact the opposite effect was found. That is, the acapsular pneumococcal variant studied showed lower hyaluronate lyase activity than the parental strain.

Since the genetic background of this variant is unknown, and nature of the mutation which lead to its phenotype of this strain is undefined, it is not possible to infer too much from this observation.

4.7 – Completion of the gene encoding hyaluronate lyase and closure of two of the pneumococcal genome contigs

As mentioned in chapter 1, section 1.8.6, the *hyl* gene sequence presented by Berry *et al.* is inconclusive, and completion of the gene sequence of the related gene from *S. agalactiae* revealed the true promoter elements preceding an N-terminal signal sequence.

Various approaches were taken to characterise the upstream region of the pneumococcal gene, which resulted in the description of an excellent promoter candidate, expression from which results in a protein that contains a predicted N-terminal signal sequence.

This observation agrees well with the various findings of the preceding chapters, including the appearance of hyaluronate lyase in the culture supernatant of a strain which lacks the C terminal cell wall anchor and the ability of pneumococci to utilise exogenous HA as a sole carbon source.

The observation that *hyl* lies 3' to the pneumococcal *gpx* gene enabled the linkage if two of the contigs of the unfinished pneumococcal genome at http://www.wit.mcs.anl.gov. This placed the locus encoding hyaluronate lyase in the context of several genes that may be involved in the utilisation of hyaluronate derived metabolites, described above. Also, it supports the proposed description of the *hyl* promoter on the basis of the lack of any other regions resembling promoter elements when reading into contig 4154.

This work also enabled the completion of the *gpx* open reading frame, as determined on the basis of likely promoter elements and homology of the proposed translation product to homologous sequences from other bacteria.

4.8 – Concluding remarks and future directions

Work described in this thesis has shown that hyaluronate lyase undergoes restricted allelic polymorphism, and is a well conserved protein. This suggests that it

is a possible candidate for trials involving the next generation of pneumococcal vaccines.

The broad range of phenotypes seen with a large group of isolates throws the broad association of hyaluronate lyase phenotype with disease-state into question, and the observation of induction of hyaluronate lyase stresses the need to observe such phenotypes under the appropriate experimental conditions.

Growth of pneumococci on HA indicates that the metabolic capability of the organism to persist on this substrate may form one pneumococcal survival strategy. The proposed role of the cellobiose-like transport system might be investigated following an interposon mutagenesis protocol of the type used in the generation of the Hyl^- strain described in Chapter 3, section 3.3.

Completion of the sequence encoding the protein reveals that it is likely to be transported to the cell surface as a result of a cleavable N-terminal signal sequence. The presence of a C-terminal anchor motif indicates that it remains associated with the cell for at least part of the time. In other streptococci (*S. intermedius*, for example) hyaluronate lyase activity is high in the cell supernatant fraction (Homer *et al.*, 1994). As such, the question of subsequent cleavage at the C-terminal portion and release into the extracellular milleu represents an interesting line of enquiry.

Alberti, S., Ashbaugh, C.D., and Wessels, M.R. (1998). Structure of the *has* operon promoter and regulation of hyaluronic acid capsule expression in group A streptococcus. *Molecular Microbiology*, **28**, 343-353.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1989). *Molecular Biology of the Cell*. New York: Garland Publishing Inc.

Alloing, G., Martin, B., Granadel, C. and Claverys, J.P. (1998). Development of competence in *Streptococcus pneumoniae*: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Molecular Microbiology*, **29**, 75-83.

AlonsoDeVelasco, E., Verheul, A.F.M., Verhoef, J. and Snippe, H. (1995). *Streptococcus pneumoniae*: Virulence factors, pathogenesis and vaccines. *Microbiological Reviews*, **59**, 591-603.

Arrecubieta, C., Garcia, E., and Lopez, R. (1995). Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. *Gene*, **167**, 1-7.

Arrecubieta, C., Lopez, R., and Garcia, E. (1996). Type 3 specific synthase of *Streptococcus pneumoniae* (Cap3B) directs type 3 polysaccharide biosynthesis in Escherichia coli and in pneumococcal strains of different serotypes. *Journal of Experimental Medicine*, **184**, (449-455).

Auzat, I., Chapuy-Regaud, S., LeBras, G., DosSantos, D., Ogunniyi, A.D., LeThomas, I., Garel, J.R., Paton, J.C. and Trombe, M.C. (1999). The NADH oxidase of *Streptococcus pneumoniae*: its involvement in competence and virulence. *Molecular Microbiology*, **34**, 1018-1028.

Avery, O., MacLeod, C.M. and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *Journal of Experimental Medicine*, **79**, 137-158.

Baker, J.R., Yu, H., Morrison, K., Averett, W.F. and Pritchard, D.G. (1997). Specificity of the hyaluronate lyase of group-B streptococcus toward unsulphated regions of chondroitin sulphate. *Biochemical Journal*, **327**, 65-71.

Benchetrit, L.C., Pahuja, S.L., Gray, E.D. and Edstrom, R.D. (1977). A sensitive method for the assay of hyaluronidase activity. *Analytical Biochemistry*, **79**, 431-437.

Benchetrit, L.C., Gray, E., Edstrom, R.D. and Wannamaker, L.W. (1978). Purification and characterisation of a hyaluronidase associated with a temperate bacteriophage of group A, type 49 streptococci. *Jounal of Bacteriology*, **134**, 221-228.

Benton, K.A., Paton, J.C. and Briles, D.E. (1997). The haemolytic and complement activating properties of pneumolysin do not contribute individually to virulence in a pneumococcal bacteraemia model. *Microbial Pathogenesis*, **23**, 201-209.

Berry, A.M., Paton, J.C., Glare, E.M., Hansman, D. and Catcheside, D.E.A. (1988). Cloning and expression of the pneumococcal neuraminidase gene in *Escherichia coli*. *Gene*, **71**, 299-305.

Berry, A.M., Lock, R.A., Hansman, D and Paton, J.C. (1989). Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infection and Immunity*, **57**, 2324-2330.

Berry, A.M., Lock, R.A., Thomas, S., Rajan, D.P., Hansman, D. and Paton, J.C. (1994). Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant *Escherichia coli*. *Infection and Immunity*, **62**, 1101-1108.

Berry, A.M. and Paton, J.C. (2000). Additive attenuation of virulence genes of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infection and Immunity*, **68**, 133-140.

Brock, T.D. (1988). *Robert Koch: A life in medicine and bacteriology*. Madison, Wisconsin: Science Tech publishers.

Brooks-Walter, A., Briles, D.E. and Hollingshead, S.K. (1999). The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicts cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteraemia. *Infection and Immunity*, **67**, 6533-6542.

Bulloch, W. (1960). *A history of bacteriology*. London: Oxford university press.

Camara, M., Mitchell, T.J., Andrew, P.W. and Boulnois, G.J. (1991). *Streptococcus pneumoniae* produces at least two distinct enzymes with neuraminidase activity: Cloning and expression of a second neuraminidase gene in *Escherichia coli*. *Infection and Immunity*, **59**, 2856-2858.

Camara, M., Boulnois, G.J., Andrew, P.W. and Mitchell, T.J. (1994). A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. *Infection and Immunity*, **62**, 3688-3695.

Camerini-Otero, R.D. and Hsieh, P. (1995). Homologous recombination proteins in prokaryotes and eukaryotes. *Annual Review of Genetics*, **29**, 509-552.

Castaneda, E., Penuela, I., Vela, M.C. and Tomasz, A. (1998). Penicillin resistant *Streptococcus pneumoniae* in Colombia: Presence of international epidemic clones. *Microbial Drug Resistance - Mechanisms, Epidemiology and Disease*, **4**, 233-239.

Cevallos, M.A., Navarro-Duque, C., Varela-Julia, M. and Alagon, A.C. (1992). Molecular mass determination and assay of venom hyaluronidases by sodium dodecyl sulphate-polyacrylamide gel electrophoresis. *Toxicon*, **30**, 925-930.

Cheng, Q., Campbell, E.A., Naughton, A.M., Johnson, S and Masure, H.R. (1997). The com locus controls genetic transformation in *Streptococcus pneumoniae*. *Molecular Microbiology*, **23**, 683-692.

Claverys, J.P., Prudhomme, M., Mortier-Barriere, I. and Martin, B. (2000). Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? *Molecular Microbiology*, **35**, 251-259.

Coffey, T.J., Dowson, C.G., Daniels, M., Zhou, J., Martin, C., Spratt, B.G. and Musser, J.M. (1991). Horizontal transfer of multiple penicillin-binding protein genes and capsular biosynthetic genes in natural populations of *Streptococcus pneumoniae*. *Molecular Microbiology*, **5**, 2255-2260.

Coffey, T.J., Dowson, C.G., Daniels, M., Spratt, B.G. (1993). Horizontal spread of an altered penicillin-binding protein 2B between *Streptococcus pneumoniae* and *Streptococcus oralis*. *FEMS Microbiology Letters*, **110**: 335-340.

Coffey, T.J., Berron, S., Daniels, M, GarciaLeoni, M.E., Cercenado, E., Bouza, E., Fenoll, A. and Spratt, B.G. (1996). Multiply antibiotic resistant *Streptococcus pneumoniae* recovered from Spanish hospitals (1988-1994): Novel major clones of serotypes 14, 19F and 15F. *Microbiology – UK*, **142**, 2747-2757.

Coffey, T.J., Enright, M.C., Daniels, M., Morona, J.K., Morona, R., Hryniewicz, W., Paton, J.C. and Spratt, B.G. (1998). Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Molecular Microbiology*, **27**, 73-83.

Coffey, T.J., Daniels, M., Enright, M.C. and Spratt, B.G. (1999). Serotype 14 variants of the Spanish penicillin resistant serotype 9V clone of *Streptococcus pneumoniae* arose by large recombinational replacements of the cpsA-pbp1a region. *Microbiology – UK*, **145**, 2023-2031.

Costagliola, C., DelPrete, A., Winkler, N.R., Carpineto, P., Ciancaglili, M., Piccolomini, R. and Mastropasqua, L. (1996). The ability of bacteria to use Na-hyaluronate as a nutrient. *ACTA Opthalmologica Scandinavica*, **74**, 566-568.

Dintilhac, A., Alloing, G., Granadel, C. and Claverys, J.P. (1997). Competence

and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Molecular Microbiology*, **25**: 727-739.

Doherty, N.C., Trzciuski, K., Pickerill, P., Zawadzki, P and Dowson, C.G. (2000). Genetic diversity of the *tet(M)* gene in tetracycline resistant clonal lineages of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, In Press.

Dorfman, A. and Ott, M.L. (1947). A turbidimetric method for the assay of hyaluronidase. *Journal of Biological Chemistry*, **172**, 367-375.

Douglas, R.M., Paton, J.C., Duncan, S.J. and Hansman, D.J. (1983). Antibody response to pneumococcal vaccination in children younger than five years of age. *Journal of Infectious Diseases*, **148**, 131-137.

Dowson, C.G., Coffey, T.J., Kell, C. and Whiley, R.A. (1993). Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. *Molecular Microbiology*, **9**, 635-643.

EchanizAviles, G., VelazquezMeza, M.E., CarnallaBarajas, M.N., SotoNogueron, A., DiFabio, J.L., SolorzanoSantos, F., JimenezTapia, Y. and Tomasz, A. (1998). Predominance of the multiresistant 23F international clone of *Streptococcus pneumoniae* among isolates from Mexico. *Microbial Drug Resistance - Mechanisms, Epidemiology and Disease*, **4**, 241-246.

Enright, M. and Spratt, B.G. (1998). A multilocus sequence typing scheme for *Streptococcus pneumoniae* : identification of clones associated with serious invasive disease. *Microbiology – UK*, **144**, 3049-3060.

Enright, M., Fenoll, A., Griffiths, D. and Spratt, B.G. (1999). The three major Spanish clones of penicillin resistant *Streptococcus pneumoniae* are the most common clones recovered in recent cases of meningits in Spain. *Journal of Clinical Microbiology*, **37**, 3210-3216.

Feinberg, R.N. and Beebe, D.C. (1983). Hyaluronate in vasculogenesis. *Science*, **220**, 1177-1179.

Ferrante, A., Rowan-Kelly, B. and Paton, J.C. (1984). Inhibition of *in vitro* human lymphocyte response by the pneumococcal toxin pneumolysin. *Infection and Immunity* **46**, 585-589.

Ferroni, A., Nguyen L., Gehanno, P., Boucot, I. and Berche, P. (1996). Clonal distribution of penicillin resistant *Streptococcus pneumoniae* 23F in France. *Journal of Clinical Microbiology*, **34**, 2707-2712.

Fethiere, J., Eggimann, B. and Cygler, M. (1999). Crystal structure of chondroitin AC lyase, a representative of a family of glycosaminoglycan degrading enzymes. *Journal of Molecular Biology*, **288**, 635-641.

Foster, W. (1970). *A History of Medical Bacteriology and Immunology*. London: Cox and Wyman Ltd.

Fox, M.S. and Hotchkiss, R.D. (1957). Initiation of bacterial transformation. *Nature*, **179**, 1322-1325.

Frost, G.I., Csoka, T. and Stern, R. (1996). The hyaluronidases: A chemical, biological and clinical overview. *Trends in Glycoscience and Glycotechnology*, **8**, 419-434.

Garcia, P., Garcia, J.L., Garcia, E. and Lopez, R. (1986). Nucleotide sequence and expression of the pneumococcal autolysin gene from its own promoter in *Escherichia coli*. *Gene*, **43**, 265-272.

Garcia, P., Paz-Gonzalez, M., Garcia, E., Lopez, R. and Garcia, J.L. (1988). LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Molecular Microbiology*, **31**, 1275-1281.

Garcia, P., Paz-Gonzalez, M., Garcia, E., Garcia, J.L. and Lopez, R. (1999). The molecular characterisation of the first autolytic lysozyme of *Streptococcus pneumoniae* reveals evolutionary mobile domains. *Molecular Microbiology*, **33**, 128-138.

Gase, K., Ozegowski, J. and Malke, H. (1998). The *Streptococcus agalactiae hylB* gene encoding hyaluronate lyase: Completion of the sequence and expression analysis.*Biochemica et Biophysica ACTA - Gene structure and expression*, **1398**, 86-98.

Giammarinaro, P., Sicard, M. and Gasc, A.M. (1999). Genetic and physiological studies of the CiaH-CiaR two-component signal-transducing system involved in cefotaxime resistance and competence of *Streptococcus pneumoniae*. *Microbiology – UK*, **145**, 1859-1869.

Gilbert, P., Allison, D.G. and Evans, D.J. (1989). Growth rate control of adherent populations. *Applied and Environmental Microbiology*, **55**, 1308-1311.

Golden, R. (1992). Oslers legacy: the centenial of the principles and practice of medicine. *Annals of International Medicine*, **116**, 225-260.

Gray, B.M. (2000). *Streptococcus pneumoniae* infections. In *Streptococcal Infections*, pp. 302-332. Edited by D.L. Stephens and E.L. Kaplan. Oxford: Oxford University Press.

Griffith, F. (1928). The significance of pneumococcal types. *Journal of Hygiene*, **27**, 113-159.

Gunther, E., Ozegowski, J-H. and Kohler, W. (1996). Occurrence of hyaluronic acid and hyaluronatlyase in streptococci of groups A, B, C and G. *International Journal of Medical Microbiology, Virology , Parasitology and Infectious Diseases*, **285**, 64-73.

Hass, E. (1946). On the mechanism of invasion. *Journal of Biological Chemistry*, **163**, 63.

Haverstein, L.S., Coomaraswarmy, G. and Morrison, D.A. (1995). An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumonie*. *PNAS*, **92**, 11140-11144.

Hodgson, S.M., Nelson, S.M., Brown, M.R.W. and Gilbert, P. (1995). A simple in vitro model for growth control of bacterial biofilms. *Journal of Applied Microbiology*, **79**, 87-93.

Hoefnagels-Schuermans, A., VanEldere, J., VanLierde, S., Verbist, L., Verhaegen, J. and Peetermans, W.E. (1999). Increase in penicillin resistance rates in Belgium due to clonal spread of a penicillin resistant 23F *Streptococcus pneumoniae strain. European Journal of Clinical Microbiology and Infectious Diseases*, **18**, 120-125.

Homer, K.A., Patel, R. and Beighton, D. (1993). Effects of N-acetylglucosamine on carbohydrate fermentation by *Streptococcus mutans* NCTC 10449 and *Streptococcus sobrinus* SL-1. *Infection and Immunity*, **61**, 295-302.

Homer, K.A., Grootveld, M.C., Hawkes, J., Naughton, D.P. and Beighton, D. (1994). Degradation of hyaluronate by *Streptococcus intermedius* strain UNS 35. *Microbial Pathogenicity*, **41**, 414-422.

Homer, K.A., Kelley, S., Hawkes, J., Beighton, D. and Grootveld, M.C. (1996). Metabolism of glycoprotein-derived sialic acid and N-acetylglucosamine by *Streptococcus oralis*. *Microbiology – UK*, **142**, 1221-1230.

Homer, K., Shain, H. and Beighton, D. (1997). Role of hyaluronidase in growth of *Streptococcus intermedius* on hyaluronate. In *Streptococci and the host*. Edited by Horaud. New York: Plenum Press.

Houldsworth, S., Andrew, P.W. and Mitchell, T.J. (1994). Pneumolysin stimulates production of tumour necrosis facor alpha and interleukin-1 beta by human mononuclear phagocytes. *Infection and Immunity*, **62**, 1501-1503.

Humbert, O., Prudhomme, M., Hakenback, R., Dowson, C.G. and Claverys, J.P. (1995). Homeologous recombination and mismatch repair during transformation in *Stretococcus pneumoniae*: Saturation of the Hex mismatch repair system. *PNAS*, **92**, 9052-9056.

Hynes, W.L. and Ferretti, J.J. (1989). Sequence analysis and expression in *Escherichia coli* of the hyaluronidase gene of *Streptococcus pyogenes* bacteriophage H4489A. *Infection and Immunity*, **57**, 533-539.

Hynes, W.L., Hancock, L. and Ferretti, J.J. (1995). Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: Evidence for a third hyaluronidase involved in extracellular activity. *Infection and Immunity*, **63**, 3015-3020.

Iannelli, F., Pearce, B.J. and Pozzi, G. (1999). The type 2 capsule locus of *Streptococcus pneumoniae*. *Journal of Bacteriology*, **181**, 2652-2654.

Ip, M., Lyon, D.J., Yung, R.W.H., Chan, C.L. and Cheng, A.F.B. (1999). Evidence of clonal dissemination of multidrug resistant Streptococcus pneumoniae in Hong Kong. *Journal of Clinical Microbiology*, **37**, 2834-2839.

Jedrzejas, M.J., Chantalat, L. and Mewbourne, R.B. (1998). Crystallisation and preliminary X-ray analysis of *Streptococcus pneumoniae* hyaluronate lyase. *Journal of Structural Biology*, **121**, 73-75.

Kass, E. and Seastone C. (1944). The role of the mucoid polysaccharide (hyaluronic acid) in the virulence of group A haemolytic streptococci. *Journal of Experimental Medicine*, **79**, 319.

Kless, H. and Vermass, W. (1995). Tandem sequence duplications functionally complement deletions in the DI protein of photosystem II. *Journal of Biological Chemistry*, **270**, 16536-16541.

Kolkman, M.A.B., Wakarchuk, W., Nuijten, P.J.M. and VanderZeijst, B.A.M. (1997). Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: Molecular analysis of the complete cps locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. *Molecular Microbiology*, **56**, 197-208.

Kostyukova, N.N., Volkova, M.O., Ivanova, V.V. and Kvetnaya, A.S. (1995). A study of pathogenic factors of *Streptococcus pneumoniae* strains causing meningitis. *FEMS Immunology and Medical Microbiology*, **10**, 133-138.

Krell, G. (1995). Hyaluronidases - A neglected group of enzymes. *Protein Science*, **4**, 1666-1669.

Kujawa, M.J. and Tepperman, K. (1983). Culturing chick muscle cells on glycosaminoglycan substrates: Attachment and differentiation. *Developmental Biology*, **99**, 277-286.

Kujawa, M.J., Pechak, D.G., Fiszman, M.Y. and Caplan, A.I. (1986). Hyaluronic acid bonded to cell culture surfaces inhibits the program of myogenesis. *Developmental Biology*, **113**, 10-16.

Kumar, S., Tamura, K. and Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis, The Pennsylvania State University.

Lacks, S. and Hotchkiss, R.D. (1960). A study of the genetic material determining an enzyme activity in *pneumococcus*. *Biochemica et Biophysica ACTA*, **39**, 508-518.

Lacks, S.A., Ayalew, S., de la Campa, A.G. and Greenberg, B. (2000). Regulation of competence for genetic transformation in *Stretococcus pneumoniae*: expression of *dpnA*, a late competence gene encoding a DNA methyltransferase of the *DpnII* restriction system. *Molecular Microbiology*, **35**, 1089-1098.

Lee, C. J., Banks, S.D. and Li, J.P. (1991). Virulence, Immunity, and Vaccine Related to *Streptococcus Pneumoniae*. *Critical Reviews in Microbiology*, **18**, 89-114.

Lee, M.S. and Morrison, D.A. (1999). Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. *Journal of Bacteriology*, **181**, 5004-5016.

Lefrancois, J., Samrakandi, M.M. and Sicard, A.M. (1998). Electrotransformation and natural transformation of *Streptococcus pneumoniae*: requirement of DNA processing for recombination. *Microbiology – UK*, **144**, 3061-3068.

Leonard, B.A.B., Woischnik, M. and Podbielski, A. (1988). Production of stabilised virulence factor-negative variants by group A streptococci during stationary phase. *Infection and Immunity*, **66**, 3841-3847.

Levin, J.C. and Wessels, M.R. (1998). Identification of *csrR/csrS*, a genetic locus that regulates hyaluronic acid capsule synthesis in group A Streptococcus. *Molecular Microbiology*, **30**, 209-219.

Levinson, G. and Gutman, G.A. (1987). Slipped-strand mispairing – A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, **4**, 203-221.

Levy, G.A. and McAllan, A. (1959). The N-Acetylation and estimation of hexosamines. *Biochemical Journal*, **73**, 127-132.

Li, S., Kelly, S.J., Lamani, E., Ferraroni, M. and Jedrzejas, M.J. (2000). Structural basis of hyaluronan degradation by *Streptococcus pneumoniae* hylauronate lyase. *EMBO Journal*, **19**, 1228-1240.

Lin, B., Averett, W.F. and Pritchard, D.G. (1997). Identification of a histidine residue essential for the enzymatic activity of group B streptococcal hyaluronate lyase. *Biochemical and Biophysical Research Communications*, **231**, 379-382.

Linker, A. and Meyer, K. (1954). The production of unsaturated uronidases by bacterial hyaluronidases. *Nature*, **174**, 1192-1193.

Llull, D., Lopez, R., Garcia, E. and Munoz, R. (1998). Molecular structure of the gene cluster responsible for the synthesis of the polysaccharide capsule of *Streptococcus pneumoniae* type 33F. *Biochimica et Biophysica ACTA-Gene Structure and Expression*, **1443**, 217-224.

Llull, D., Munoz, R., Lopez, R. and Garcia, E. (1999). A single gene (tts) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide: Type 27 are natural, genetically binary strains. *Journal of Experimental Medicine*, **190**, 241-251.

Lopez, R., Sanchez-Puelles, J.M., Garcia, E., Garcia, J.L., Ronda, C. and Garcia, P. (1986). Isolation, characterisation and physiological properties of an autolytic deficient mutant of *Streptococcus pneumoniae*. *Molecular and General Genetics*, **204**, 237-242.

McClean, D. and Hale, C.W. (1941). Studies on diffusing factors: The hyaluronidase activity of testicular extracts, bacterial culture filtrates and other agents that increase tissue permeability. *Biochemical Journal*, **35**, 159.

McGee, L., Klugman, K.P., Friedland, D. and Lee, H.J. (1997). Spread of the Spanish multi-resistant serotype 23F clone of *Streptococcus pneumoniae* to Seoul, Korea. *Microbial Drug Resistance - Mechanisms, Epidemiology and Disease*, **3**, 253-257.

McGeoch, D.J. (1985). On the predictive recognition of signal peptide sequences. *Virus Research*, 3, 271-286.

Majewski, J., Zawadski, P., Pickerill, P., Cohan, F.M. and Dowson, C.G. (2000). Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *Journal of Bacteriology*, **182**, 1016-1023.

Marciel, A.M., Kapur, V., Musser, J.M. (1997). Molecular population genetic analysis of a *Streptococcus pyogenes* bacteriophage-encoded hyaluronidase gene: Recombination contributes to allelic variation. *Microbial Pathogenesis*, **22**, 209-217.

Masure, R.H., Pearce, B.J., Shio, H. and Spellerberg, B. (1998). Membrane targeting of RecA during genetic transformation. *Molecular Microbiology*, **27**, 845-852.

Maynard Smith, J. (1992). Analysing the mosaic structure of genes. *Molecular Biology and Evolution*, **34**, 126-129.

Michal, G. (1993). *Biochemical pathways*. Third edition, part 1. Distributed by Boehringer-Manheim GmbH.

Mitchell, T.J., Andrew, P.W., Saunders, F.K., Smith, A.N. and Boulnois, G.J. (1991). Complement activation and antibody binding by pneumolysin via a region of the toxin homologous to a human acute-phase protein. *Molecular Microbiology*, **5**, 1883-1888.

Mitchell, T.J., Alexander, J.E., Morgan, P.J. and Andrew, P.W. (1997). Molecular analysis of virulence factors of *Streptococcus pneumoniae*. *Journal of Applied Microbiology*, **83** (supplement), 62S-71S.

Morona, J.K., Morona, R. and Paton, J.C. (1999a). Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. *Journal of Bacteriology*, **181**, 5355-5364.

Morona, J.K., Miller, D.C., Coffey, T.J., Vindurampulle, C.J., Spratt, B.G., Morona, R. and Paton, J.C. (1999b). Molecular and genetic characterisation of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 23F. *Microbiology – UK*, **145**, 781-789.

Morrison, D.A. (1978). Transformation in pneumococcus: protein content of eclipse complex. *Journal of Bacteriology*, **136**, 548-557.

Morrison, D.A. and Mannarelli, B. (1979). Transformation in pneumococcus: nuclease resistance of deoxyribonucleic acid in the eclipse complex. *Journal of Bacteriology*, **140**, 655-665.

Morrison, D.A. (1997). Streptococcal competence for genetic transformation: regulation by peptide hormones. *Microbial Drug Resistance*, **3**, 27-37.

Mortier-Barriere, I., deSaizieu, A., Claverys, J.P. and Martin, B. (1998). Competence-specific induction of recA is required for full recombination proficiency during transformation in *Streptococcus pneumoniae*. *Molecular Microbiology*, **27**, 159-170.

Moses, A.E., Wessels, M.E., Zalcman, K., Alberti, S., Natanson-Yaron, S., Menes, T. and Hanski, E. (1997). Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A streptococcus. *Infection*

and Immunity, **65**, 64-71.

Muller-Graf, C.D.M., Whatmore, A., King, S.J., Trzcinski, K., Pickerill, A.P., Doherty, N., Paul, J., Griffiths, D., Crook, D. and Dowson, C. (1999). Population biology of *Streptococcus pneumoniae* isolated from oropharyngeal carriage and invasive disease. *Microbiology – UK*, **145**, 3283-3293.

Munoz, R., Coffey, T.J., Daniels, M., Dowson, C.G., Laible,G., Casal, J., Hakenbeck, R., Jacobs, M., Musser, J.M., Spratt, B.G. and Tomasz, A. (1991). Intercontinental spread of a multiresistant clone of serotype 23F *Streptococcus pneumoniae*. *Journal of Infectious Diseases*, **164**, 302-306.

Munoz, R., Mollerach, M., Lopez, R. and Garcia, E. (1997). Molecular organisation of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: Formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamanose biosynthesis genes. *Molecular Microbiology*, **25**, 79-92.

Munoz, R., Mollerach, M., Lopez, R. and Garcia, E. (1999). Charactersation of the type 8 capsular gene cluster of *Streptococcus pneumoniae*. *Journal of Bacteriology*, **181**, 6214-6219.

Navarre, W.W. and Schneewind, O. (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews*, **63**, 174-229.

Nei, M and Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269-5273.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, **10**, 1-6.

Novak, R., Charpentier, A., Charpentier, E and Tuomanen, E. (1999). Identification of a *Streptococcus pneumoniae* gene locus encoding proteins of an ABC phosphate transporter and a two-component regulatory system. *Journal of Bacteriology*, **181**, 1126-1133.

Paoletti, L.C., Kasper, D.L., Michon, F., DiFabio, J., Jennings, H.J., Tostenson, T.D. and Wessels, M.R. (1992). Effects of chain length on the immunogenicity in rabbits of group B streptococcus type III oligosaccharide-tetanus toxoid conjugates. *Jounal of Clinical Investigation*, **89**, 203-209.

Paton, J.C. and Ferrante, A. (1983). Inhibition of human polymorphonuclear

leukocyte respiratory burst, bactericidal activity and migration by pneumolysin. *Infection and Immunity*, **41**, 1212-1216.

Paton, J.C., Andrew, P.W., Boulnois, G.J. and Mitchell, T.J. (1993). Molecular analysis of the pathogenicity of *Streptococcus pneumoniae*: The role of pneumococcal proteins. *Annual Review of Microbiology*, **47**, 89-115.

Paton, J.C., Berry, A.M. and Lock, R.A. (1997). Molecular analysis of putative pneumococcal virulence proteins. *Microbial Drug Resistance-Mechanisms, Epidemiology and Disease*, **3**, 1-10.

Peeters, C.C.A.M., Tenbergen-Meekes, A.M., Evenberg, D.E., Poolman, J.T., Zegers, B.J.M. and Rijkers, G.T. (1991). A comparative study of the immunogenicity of pneumococcal type 4 polysaccharide and oligosaccharide tetanus toxoid conjuates in adult mice. *Journal of Immunology*, **146**, 4308-4314.

Peri, K.G. and Watgood, E.B. (1988). Sequence of cloned enzyme $II^{N\text{-acetylglucosamine}}$ of the phosphoenolpyruvate:N-acetylglucosamine phosphotransferase system of *Escherichia coli*, *Biochemistry*, **27**, 6054-6061.

Pestova, E.V., Haversiein, L.S. and Morrison, D.A. (1996). Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Molecular Microbiology*, **21**, 853-862.

Plotkowski, M.C., Pluchelle, E., Beck, G., Jaquot, J. and Hannoun, C. (1986). Adherence of type 1 *Streptococcus pneumoniae* to tracheal apithelium of mice infected with influenza A/PR8 virus. *American Review of Respiratory Diseases*, **134**, 1040-1044.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Motti, H., Ferarri, L. and Simon, D. (1998). Large-scale identification of virulence genes from *Streptococcus pneumoniae*, *Infection and Immunity*, **66**, 5620-5629.

Ponnuraj, K. and Jedrzejas, M.J. (2000). Mechanism of hyaluronan binding and degradation: Structure of *Streptococcus pneumoniae* hyaluronate lyase in complex with hyaluronic acid disaccharide at 1.7A resolution. *Journal of Molecular Biology*, **299**, 885-895.

Poulsen, K., Reinholdt, J. and Kilian, M. (1996). Characterisation of the *Streptococcus pneumoniae* immunoglobulin A1 protease gene (*iga*) and its translation product. *Infection and Immunity*, **64**, 3957-3966.

Poulsen, K., Reinholdt, J., Jespersgaard, C., Boye, K., Brown, T.A., Hauge, M. and Kilian, M. (1998). A comprehensive genetic study of streptococcal immunoglobulin A1 proteases: Evidence for recombination within and between

species. *Infection and Immunity*, **66**, 181-190.

Primakoff, P. and Myles, D.G. (1983). A map of the guinea pig sperm surface constructed with monoclonal antibodies. *Developmental Biology*, **98**, 417-428.

Pritchard, D., Lin, B., Willingham, T.R. and Baker, J.R. (1994). Characterisation of the group B *Streptococcal* hyaluronate lyase. *Archives of Biochemistry and Biophysics* **315**, 431-437.

Proctor, M. and Manning, P. (1990). Production of Immunoglobulin A protease by *Streptococcus pneumoniae* from animals. *Infection and Immunity*, **58**, 2733-2737.

Ramirez, M., Tomasz, A. (1999a). Acquisition of new capsular genes among clinical isolates of antibiotic resistant *Streptococcus pneumoniae*. *Microbial Drug Resistance - Mechanisms, Epidemiology and Disease*, **5**, 241-246.

Ramirez, M. and Tomasz, A. (1999b). Molecular characterisation of the complete 23F capsular polysaccharide locus of *Streptococcus pneumoniae*. *Journal of Bacteriology*, **180**, 5273-5278.

Ravin, A.W. (1959). Reciprocal capsular transformation of pneumococci. *Journal of Bacteriology*, **77**, 296-309.

Reed, R.K., Laurent, U.B.G., Fraser, J.R.E. and Laurent, T.C. (1990a). Removal rate of [^{3}H]-hyaluronan injected subcutaneously in rabbits. *American Journal of Physiology*, **259**, H532-H535.

Reed, R.K., Laurent, T.C. and Taylor, A.E. (1990b). Hyaluronan in prenodal lymph from the skin: Changes with lymph flow. *American Journal of Physiology*, **259**, H1097-1100.

Reinholdt, J. and Kilian, M. (1997). Comparative analysis of immunoglobulin A1 protease activity among bacteria representing different genera, species and strains. *Infection and Immunity*, **65**, 4452-4459.

Rimini, R., Jansson. B., Feger, G., Roberts, T.C., deFrancesco, M., Gozzi, A., Faggioni, F., Domenici, E., Wallace, D.M., Frandsen, N. and Polissi, A. (2000). Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Molecular Microbiology*, **36**, 1279-1292.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A. and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonisation and immunogenicity of *Streptococcus pneumoniae*. *Molecular Microbiology*, **25**, 819-829.

Rossjohn, J., Gilbert, R.J.C., Crane, D., Morgan, P.J., Mitchell, T.J., Rowe, A.J., Andrew, P.W., Paton, J.C., Tweten, R.K. and Parker, M.W. (1998). The molecular mechanism of pneumolysin, a virulence factor from *Stretococcus pneumoniae*. *Journal of Molecular Biology*, **284**, 449-461.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbour Laboratory Press.

Sanchez-Beato, A.R., Lopez, R. and Garcia, J.L. (1998). Molecular characterisation of PcpA: A novel choline-binding protein of *Streptococcus pneumoniae*. *FEMS Microbiology Letters*, **164**, 207-214.

Sato, K., Quartey, M.K., Liebeler C.L., Le, C.T. and Giebink, G.S. (1996). Roles of autolysin and pneumolysin in middle ear inflammation caused by a type 3 *Streptococcus pneumoniae* strain in the chinchilla otitis media model. *Infection and Immunity*, **64**, 1140-1145.

Sawyer, S. (1989). Statistical tests for detecting gene conversion. *Molecular Biology and Evolution*, **6**, 526-538.

Schrager, H.M., Alberti, S., Cywes, C., Dougherty, G.J. and Wessels, M.R. (1998). Hyaluronic acid capsule modulates M protein-mediated adherence and acts as a ligand for attachment of group A streptococcus to CD44 on human keratinocytes. *Journal of Clinical Investigation*, **101**, 1708-1716.

Scott, D., Siboo, I.R., Chan, E.C.S. and Siboo, R. (1996). An extracellular enzyme with hyaluronidase and chondroitinase activities from some oral anaerobic spirochaetes. *Microbiology – UK*, **142**, 2567-2576.

Seastone, C.V. (1939). The virulence of group C haemolytic streptococci of animal origin. *Journal of Experimental Medicine*, **70**, 361.

Sibold, C., Waing, C.J., Henrichsen, J. and Hackenbeck, R. (1992). Genetic relationships of penicillin-susceptible and penicillin-resistant *Streptococcus pneumoniae* strains isolated on different continents. *Infection and Immunity*, **60**, 4119-4126.

Smith, R. and Willet, N. (1968). Rapid plate method for screening hyaluronidase and chondroitin sulphatase producing microorganisms. *Applied Microbiology*, **16**, 1434-1436.

Soares, S., Kristiansson, K.G. and Musser, J.M. (1993). Evidence for the introduction of a multiresistant clone of the serotype 6B *Streptococcus pneumoniae* from Spain to Iceland in the late 1980's. *Journal of Infectious Diseases*, **168**, 158-163.

Stamnekovic, I. and Aruffo, A. (1994). Hyaluronic acid receptors. *Methods in Enzymology*, **245**, 195-216.

Steinfort, R.A., Wilson, R. and Mitchell, T. (1989). Effect of *Streptococcus pneumoniae* on human respiratory epithelium *in vitro*. *Infection and Immunity*, **57**, 2006-2013.

Sutherland, I.W. (1995). Polysaccharide lyases. *FEMS Microbiology Reviews*, **16**, 323-347.

Swaitlo, E., Brookes-Walter, A., Briles, D.E. and McDaniel, L.S. (1997). Oligonucleotides identify conserved and variable regions of *pspA* and *pspA*-like sequences of *Streptococcus pneumoniae*. *Gene*, **188**, 279-284.

Talkington, D.F., Crimmins, D.L., Voellinger, D.C., Yother, J. and Briles, D.E. (1991). A 43-kilodalton pneumococcal surface protein, PspA: Isolation, protective abilities and structural analysis of the amino terminal sequence. *Infection and Immunity*, **59**, 1285-1289.

Tomasz, A. (1965). Control of the competent state in pneumococcus by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. *Nature*, **208**, 155-159.

Tong, H.H., Blue, L.E., James, M.A. and DeMaria, T.F. (2000). Evaluation of the virulence of a *Streptococcus pneumoniae* neuraminidase-deficient mutant in nasopharyngeal colonisation and development of otitis media in the chinchilla model. *Infection and Immunity*, **68**, 921-924.

Toole, B.P. (1981). Glycosaminoglycans in Morphogenesis. In *Cell Biology of the Extracellular Matrix*. Edited by E.D. Hay. New York: Plenum Press.

Tu, A.H.T., Fulgham, R.L., McCrory, M.A., Briles, D.E. and Szalai, A.J. (1999). Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infection and Immunity*, **67**, 4720-4724.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979). Hybridisation of synthetic oligoribonucleotides to phi-chi DNA: The effect of single base pair mismatch. *Nucleic Acids Research*, **6**, 3543-3557.

Wani, J.H., Gilbert, J.V., Plaut, A.G. and Weiser, J.N. (1996). Identification, cloning and sequencing of the immunoglobulin A1 protease gene of *Streptococcus pneumoniae*. *Infection and Immunity*, **64**, 3967-3974.

Watson, D. and Musher, D.M. (1990). Interruption of capsule production in *Streptococcus pneumoniae* serotype 3 by insertion of transposonTn916. *Infection and Immunity*, **58**, 3135-3138.

Ween, O., Gaustad, P. and Havarstein, L.S. (1999). Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Molecular Microbiology*, **33**, 817-827.

Weiser, J.N. and Kapoor, M. (1999). Effect of intra-strain variation in the amount of capsular polysaccharide on genetic transformation of *Streptococcus pneumoniae*: Implications for virulence studies of encapsulated strains. *Infection and Immunity*, **67**, 3690-3692.

West, D.C., Hampson, I.N., Arnold, F. and Kumar, S. (1985). Angiogenesis induced by degradation products of hyaluronic acid. *Science*, **228**, 1324-1326.

Whatmore, A.M., King, S.J., Doherty, N.C., Sturgeon, D., Chanter, N. and Dowson, C.G. (1999). Molecular characterization of equine isolates of *Streptococcus pneumoniae*: Natural disruption of genes encoding the virulence factors pneumolysin and autolysin. *Infection and Immunity*, **67**, 2776-2782.

Whatmore, A.M. and Dowson, C.G. (1999). The autolysin-encoding gene (*lytA*) of *Streptococcus pneumoniae* displays restricted allelic variation despite localised recombination events with genes of pneumococcal bacteriophage encoding cell wall lytic enzymes. *Infection and Immunity*, **67**, 4551-4556.

Whatmore, A.M., Efstratiou, A., Pickerill, A.P., Broughton, K., Woodward, G., Sturgeon, D., George, R. and Dowson, C.G. (2000). Genetic relationships between clinical isolates of *Streptococcus pneumoniae*, *Streptococcus oralis* and *Streptococcus mitis*: Characterisation of "atypical" pneumococci and organisms allied to *S. mitis* harbouring *S. pneumoniae* virulence factor-encoding genes. *Infection and Immunity*, **68**, 1374-1382.

Winter, A.J., Comis, S.D., Osborne, M.P., Tarlow, M.J., Stephen, J., Andrew, P.W., Hill, J. and Mitchell, T.J. (1997). A role for pneumolysin but not neuraminidase in the hearing loss and cochlear damage induced by experimental pneumococcal meningitis in guinea pigs. *Infection and Immunity*, **65**, 4411-4418.

Yamagata, T., Saito, H., Habuchi, O. and Suzuki, S. (1968). Purification and properties of bacterial chondroitinases and chondrosulfatases. *Journal of Biological Chemistry*, **243**, 1523-1535.

Yanisch-Perron, C., Viera, J. and Messing, J. (1985). Improved M13 phage vectors and host strains - Nucleotide sequences of the M13MP18 and pUC19 vectors. *Gene*, **33**, 103.

Yother, J., Leopold, K., White, J. and Fischer, W. (1998). Generation and properties of a *Streptococcus pneumoniae* mutant which does not require choline or analogs for growth. *Jounal of Bacteriology*, **180**, 2093-2101.

Yuki, H. and Fishman, W.H. (1963). Purification and characterisation of Leech hyaluronic acid endo-B-glucuronidase. *Journal of Biological Chemistry*, **238**, 1877-1879.